

GLOBALIZATION: PAST, PRESENT AND FUTURE

I Globalization in Historical Perspective (up to 1990)

**by
Prof. Dr. Karl-Heinz Paqué**

**Otto-von-Guericke-University Magdeburg
Winter Term 2018/19**

**This manuscript is for internal use of students only. Quotations or
any other use must be cleared with the author (paque@ovgu.de).**

CONTENTS

	Page
1 Introduction	003
1.1 Elements of Globalization	004
1.1.1 Trade in Goods and Services	004
1.1.2 Labour Mobility	006
1.1.3 Capital Mobility	007
1.1.4 Diffusion of Technology	010
1.2 Consequences of Globalization	011
1.2.1 Economic Growth	011
1.2.2 Cyclical Fluctuations	012
1.2.3 Income Distribution	013
2 Early Trends of Integration	014
2.1 The Ancient World of Trade	016
2.1.1 Phoenicians and Greeks	016
2.1.2 Roman Empire	019
2.1.3 Sliding into a Dark Age	029
2.2 Middle Ages and Renaissance	032
2.2.1 Early Feudalism	032
2.2.2 The Great Expansion 1000-1350	040
2.2.2.1 Urbanisation	042
2.2.2.2 Technical Progress	048
2.2.3 The Time of Scarce Labour 1350-1500	051
2.2.4 Maritime Trade	056
2.2.4.1 Southern Europe: Venice and Genoa	059
2.2.4.2 Northern Europe: The Hanseatic League	067
2.2.5 Land Trade	071
2.3 Towards the Modern Age (1500-1820)	072
2.3.1 Europe's Overseas Expansion	074
2.3.1.1 Portugal and Spain	074
2.3.1.2 Holland and England	076
2.3.2 Scientific Advances	080
2.3.3 An Energy Crisis	082
2.3.4 The Infancy of Modern Industry	084
3 Globalization 1820 - 1990	085
3.1 The First Era of Integration: 1820-1914	085
3.1.1 Industrialization	089
3.1.1.1 Britain's Lead 1820-50	100
3.1.1.2 Others' Catching Up 1850-1914	102
3.1.2 R&D-based Industrialization	105
3.1.3 Steamships, Railways and Telecommunication	106
3.1.4 International Trade	108

3.1.4.1	The Path to Free Trade 1820-1878	109
3.1.4.2	Protectionist Tendencies 1878-1914	120
3.1.5	International Capital Movements	124
3.1.5.1	The Gold Standard	127
3.1.5.2	The Role of London	130
3.1.6	International Migration	132
3.2	<u>The Era of Disintegration: 1914-1950</u>	137
3.2.1	World War I and its Aftermath	138
3.2.2	The Great Inflation 1922/23	141
3.2.3	Temporary Stabilisation	144
3.2.4	The Great Depression 1930/32	146
3.2.5	Paths towards Autarky	157
3.2.6	World War II and its Aftermath	158
3.3	<u>The Second Era of Integration: 1950-90</u>	159
3.3.1	Rapid Post-War Growth 1950-73	159
3.3.1.1	The American Lead	161
3.3.1.2	Others' Catching up	161
3.3.1.3	The Surge of Intra-Industry-Trade	165
3.3.1.4	The Bretton Woods System	167
3.3.2	Oil Crises and their Aftermath 1973-90	171

1 Introduction

Everybody speaks about Globalization these days. In this lecture, we do so too. But we do it in a much more precise way than is usually done in the public. This is why we first clarify what we mean by Globalization. Without indulging in extensive debates on definitions, we briefly explain how we delimit the term “Globalization” for the purpose of this lecture. In this respect, three points stand out.

Firstly, we focus on the **economics of Globalization**. To be covered in this lecture, an aspect of Globalization must have something to do with the world-wide allocation of resources and/or with the growth (or shrinkage) of the world economy or relevant parts of it. Of course, the dividing line between the economic and the non-economic spheres of Globalization is less clear-cut than we would like it to be. As a general guideline, however, we will only cover non-economic aspects if they are indispensable for the understanding of the economic phenomena in question. As we shall see, this is quite often the case for political, but sometimes also for social and cultural aspects of the matter.

Secondly, we look at **Globalization all over history**, not only for most recent times. While it is true that the term Globalization has become widespread not before the late 1990s, the phenomenon – if properly understood – is much older. Powerful processes of economic Globalization can be found as early as the 19th century, and important single elements of it go as far back as early modern times, the Middle Ages and even the ancient world. In fact, one may argue with some legitimacy that it is hardly possible to understand the present-day trends without getting at least a brief summary exposure to what happened in earlier decades and centuries. In this lecture, we try to deliver just this exposure – before moving on to a more detailed account of recent developments. Note that, despite this important historical dimension, the lecture is a course in economics, not in economic history. We take the results of historical research as given. Or, to put it bluntly: we want to interpret the facts economically; we do not want to check whether the facts are true or false.

Thirdly, when analyzing Globalization, one has to make a **choice of regions** that are to be in the focus. E. g., Globalization in the 19th century cannot be understood without a thorough look at Britain, the most highly industrialised economy and the dominant trading nation of the time. The same holds, further back in history, for Holland in the 17th or Italy in the 14th century; or, close to the present, for the United States and Western Europe in the second half of the 20th century; or for China and India in the most recent past. All in all, this choice of focus is unavoidable and non-arbitrary, but it may introduce an undue bias into the general per-

spective. This has to be kept in mind in the course of the analysis. We shall return to the danger of biased judgements at various points of the lecture.

So much for the scope and limits of our subject “Globalization” – and, for that matter, of the second, more general part of the lecture’s title: “world economic development”. Let us now turn to what we call the **elements of Globalization**. Obviously, Globalization has something to do with a “flow of something” across national or regional borders and boundaries. What is this “something” that flows? Using standard economic reasoning (think, e. g., of an aggregate production function), four major flows come to mind: (i) trade in goods and services, (ii) migration of labour, (iii) movements of capital and (iv) diffusion of technological ideas. For the rest of this introduction, we shall briefly consider each of these flows separately (section 1.1) and then analyze the likely **consequences of Globalization** due to these flows in terms of economic growth, cyclical fluctuations and the distribution of income among and within regions or countries (section 1.2).

1.1 Elements of Globalization

1.1.1 Trade in Goods and Services

Trade in goods and services is certainly the most obvious and fundamental element of Globalization. Nowadays, trade is mostly conceived as **international** trade, i. e. as that part of trade in goods and services that crosses national borders. This view has its apparent limits: national borders are first of all legal entities, not economic ones. They gain economic significance only to the extent that they coincide with economically relevant trade barriers such as tariffs or national differences in business practices, administration, language etc. An isolated analysis of international trade at the exclusion of intra-national trade between regions or local units may thus seriously distort the economic message. E. g., an expansion of (international) trade between Bavaria and neighbouring Austria may run parallel to an expansion of (intra-national, but inter-state) trade between Bavaria and neighbouring Baden-Württemberg; in fact, both may be the outcome of a common economic cause and may have the same economic consequences so that they should be part of a common story of “Globalization”.

In historical terms, it is even more obvious that we cannot focus on international trade alone. The reason is simple: nations in the modern sense do not really exist for a very long time, and one would have lots of trouble to understand early phenomena of trade integration if one were to define it in terms of national borders and boundaries. E. g., in the Middle Ages, cities

of northern Italy and the German Hanseatic League traded among themselves and with the rest of the world. Was that international trade? And if not, what was it? Obviously, it is much more sensible to look for other more economic categories of trade that may be useful for particular purposes. E. g., in the Middle Ages, a period of history that was generally characterized by high transport costs and high risks of travel, distinctions between long distance trade and short distance trade or between maritime trade and land trade may be much more helpful to get some reasonable structure into the analysis.

In economics, there is a body of theory that explains the causes and consequences of trade in goods and services: the theory of international trade. Here again, we hit upon the term “international” although the theory, properly interpreted, is a general theory of trade, and not just a theory of international trade. Historically, of course, the theory was born and developed with a focus on the international part of trade, which inspired virtually all great contributions to theory ever since the two classical works, Adam Smith’s “The Wealth of Nations” published in 1776 and David Ricardo’s “Principles of Political Economy and Taxation” published in 1817. Nevertheless, there is no inherent reason why trade theory should be confined to trade among nations although, of course, it is the international background that gives the theory its most fascinating touch.

For the purpose of this lecture, a good knowledge of international trade theory is a big help, though not an absolute precondition for an understanding of the subject matter. To facilitate this understanding, we shall give the relevant theoretical background and interpretation on the spot, i. e. when we come to the phenomena in question. We will touch upon all major branches of trade theory because, at different times in the history of Globalization, one needs different strands of theory to interpret the facts.

At this point, it suffices to say that, on a most basic level, voluntary trade has one cause and one consequence:

The most basic cause of trade is **profit**. Trade of goods and services will only occur on a voluntary basis if those engaged in it expect to gain something from it. This is trivial, but important. It implies, e. g., that, whenever transport costs are very high for whatever reason, it will not pay off to engage in trade, at least not over longer distances. As we shall see, vast parts of economic history and Globalization up to most recent times can be told in terms of a decline in transport and communication costs and the reduction of political trade barriers, which enlarge the leeway for profitable trade.

The most basic consequence of trade is **specialisation**. Once it takes place, trade leads to a change in the division of labour: “exported” goods will be produced more and “imported” goods less than without trade. And this, as a rule, is economically efficient for all countries involved because the pattern of specialisation moves towards what everybody is best at doing. This – of course, expressed in more precise terms of economic opportunity costs – is the essence of the famous “law of comparative advantage”, which will pop up repeatedly in the course of the lecture. New patterns of specialisation, in turn, can have important repercussions on the productivity of labour and of capital, and thus also on factor prices, i. e. wages for labour and rates of return on capital. In fact, the gains from trade may be allocated unevenly, which makes things politically complicated. As we shall see, economic history is full of fascinating examples of major effects of trade on productivity growth and the distribution of income in the economy.

1.1.2 Labour Mobility

Labour is mobile. It has always been so, all over human history. People tend to move to those places where they have the best economic prospects for themselves and their families. They move within and between countries. They move on a temporary basis, expecting to return after the season or a cyclical boom, and possibly sending remittances home. Or they move forever – thus becoming immigrants in the region or country of destination. In some periods of history, these movements grew into dramatic dimensions, most notably in the 19th century when large waves of emigrants left Europe across the Atlantic to take advantage of the ample supply of land in North America. In other periods, the extent of people’s migration remained moderate.

By historical standards, today’s level of worldwide labour migration is not very large. However, this is mainly due to fairly high barriers to immigration that have been erected around the rich parts of the world. Without these, there would almost certainly be vast movements of labour on a world-wide scale. In this sense, there is nowadays a huge “pent-up demand” for migration: given low costs of transport and communication, people would like to move from poor to rich countries, but they are not allowed to do so.

Note that, historically, this situation is far from normal. There have been numerous periods in history when immigrants were highly welcome in the country or region of destination because labour was scarce and more labour was needed to expand production and generate higher incomes. In particular, skilled labour was welcome: when migrating, a worker does not only carry his physical force with him, but also his skills and his brain power. In a sense, he has a

stock of capital attached to him - in economic jargon: "human capital" -, and this human capital may be complementary to other factors of production such as land, "raw" labour and physical capital in the country of destination.

As in the case of trade, we leave the more detailed analysis of the causes and consequences of labour mobility to later parts of this lecture when we look at specific trends of Globalization in the past and present. On a very basic level, however, the economics of labour migration is straightforward and simple: in the country of destination, migration increases the labour supply and thus lowers the wage level (other things being equal); in the country of origin, it decreases the labour supply and thus raises the wage level (again, other things being equal). If migration runs from a low-wage to a high-wage country, which is the normal case, then it works in the direction of an international equalisation of wages.

Once elements of human capital enter the picture, things get somewhat more complex. An increase in the supply of capital – be it human or physical – increases the capital stock that is complementary to "raw" labour, and thus raises labour productivity and the wage level in the country of destination. The reverse holds for the country of origin. Hence the migration of human capital from poor to rich countries tends to widen the wage gap between rich and poor because it works like a drain on badly needed complementary resources of the poor country.

1.1.3 Capital Mobility

Savings can be invested at home or abroad. This is why capital is mobile, within and between countries. Remarkably enough, it is the mobility of capital across national borders that pushed Globalization into the limelight of the public's interest in recent years, much more so than labour mobility and even trade integration. As we shall see in Part 4 of the lecture, this is mainly due to the drastic decline of communication costs and the liberalisation of financial markets, which both took place in the course of the most recent decades. In spite of this, it is important to note that, just like labour mobility, capital mobility across regions and nations is not something new, but has a long history that goes back to the 19th century and even much earlier times.

To understand capital mobility, it is useful to make a distinction between "portfolio investment" on one side and "direct investment" on the other:

- **Portfolio investments** are defined as those capital flows that are purely financial in the sense that the investor's aim is to achieve a satisfactory rate of return, with no entrepreneurial or strategic side goals whatsoever. Portfolio investments cover a vast range of financial assets: from current and saving accounts in banks up to fixed-interest bonds and modern financial market derivatives. They may also cover shares in private firms provided that there is no strategic aim behind the shareholding. Note that, by themselves, portfolio investments do not increase the capital stock in the physical sense because they do not enlarge the productive capacity of an economy; rather they increase asset demand and thus lead to a revaluation of the capital stock that already exists in the country of destination.
- **Direct investments** are defined as those capital flows that are part or follow from a strategic entrepreneurial decision to invest in a specific location or region, be that within or outside a country. In the latter case, we speak of foreign direct investment (FDI). Thus defined, direct investments are just like a normal investment by a firm owner or manager in plant and equipment that has just one distinguishing characteristic: the investment is not made "at home", i. e. in the traditional location, but "abroad" – whatever this means in terms of geography. Note that direct investments do increase the capital stock in the country of destination if (and only if) they do lead to an expansion of productive capacity (e.g., the building of a new plant or the enlargement of an old one). This may be, but it need not be the case: a strategically motivated purchase of shares of an existing foreign firm at the stock exchange qualifies as a direct investment, but does not yet imply an expansion of productive capacity.

Clearly, both types of capital flows are major constituent elements of a Globalization in capital markets. Both have gained in quantitative importance in modern times though it is worth noting that the speed of expansion of portfolio investments has vastly outpaced that of direct investments. Note that, in practice, there are many hybrid forms between them, which may make a sharp categorization difficult.

As to the causes of capital mobility, it is again useful to distinguish between portfolio and direct investments:

- **Portfolio investments** are driven by a straight profit motive. Capital flows to those places where the rate of return on investment is highest, with an appropriate consideration of risks. Naturally, any rate of return calculation can be highly complex and speculative: from international differentials in nominal interest rates over the monetary

equivalent of default risks down to exchange rates of currencies and their expected changes. In the end, however, it is the rate of return that counts, and hardly anything else.

- **Direct investments** are driven by a much more complex set of motives. As they are part of a more general strategic and entrepreneurial planning, they have to take all parameters into account that are typically part of this planning, e. g., production costs, the regional or national market size, growth prospects and the state of the infrastructure. Of course, in the end, there must be a sufficient rate of return on capital to justify a direct investment, but the link between the economic decision and the profit motive is much more indirect than in the case of portfolio investments.

As to the consequences of capital mobility, it is safe to say that, in general, a flow of capital from a country with a low to a country with a high rate of return leads towards an equalisation of rates of return. If capital markets work properly, this means that, just as rates of return, so the degrees of capital scarcity tend to be equalized: the country that most urgently needs capital imports it, and the country that least urgently needs capital exports it. Of course, the real world may be much more complex: in recent years, e. g., relatively poor China has been massively exporting capital to the relatively rich United States, and that is not easily explained in terms of whatever comparative capital scarcity of nations.

Naturally, there may be other and even more fundamental complications, which we can only briefly mention at this point with the promise to return to them in the course of the lecture.

- In the case of **portfolio investments**, waves of speculation may distort the market signals. These may lead to tides of capital flows in and out of a country and thus to short-term revaluations of national or regional capital stocks that have only loose links to fundamental economic data on (long-term) scarcities. In fact, such short-term waves and tides may have far-reaching macroeconomic effects that we will deal with in various parts of the lecture, notably in part 3 of this manuscript when we discuss the Great Depression 1930/32 and other big world-wide recessions. Note also that, at present in 2007/8, we are going through a financial crisis that may possibly attain world-wide dimensions with still unknown macroeconomic consequences (the so-called US subprime crisis).
- In the case of **direct investments**, the link between the scarcity of capital and the profitability of capital is a very loose one anyway. In fact, there is no compelling rea-

son why the strategic decision of some entrepreneur to invest in a new plant abroad – and not at home – should have anything to do with the macroeconomic states of capital scarcity at home and abroad. Of course, in the country of destination, an increase in the supply of capital has salutary effects on productive capacity because it endows labour with a new capital stock and thus raises labour productivity and the wage level in the country of destination. However, the reverse holds for the country of origin, and – a priori – the net balance is open.

1.1.4 Diffusion of Technology

How can technical knowledge cross-regional and/or national boundaries? The obvious answer is: as a companion to something else that flows across borders. That something may be goods and services that embody technical knowledge, which can then be analyzed by experts in the importing country. It may also be the knowledge and skills that an immigrant worker imports when moving from his old to his new home country (his “human capital”). And it may be the knowledge incorporated in sophisticated machinery that is installed and serviced by local labour in the country that receives a foreign direct investment.

These three variants of knowledge diffusion constitute the set of standard cases. As we shall see in the course of the lecture, they do in fact play a very important – and over time increasing – role in the economic history of Globalization. After all, it is normal that knowledge comes attached to a human being or a physical object. Yet there are other important variants of knowledge diffusion, which have a “disembodied” character. Some of them come to mind immediately:

- Firms may buy licences to use foreign patents in their production process. They thus import – at a price – an established piece of protected technology from abroad, a classical case of “buying knowledge”.
- Students (or apprentices) may study or learn abroad and import technical knowledge after their return home. The knowledge transfer then has the character of a direct service import of education.
- People may read foreign books or browse the internet, thus directly importing technical information from abroad, a kind of knowledge transfer that is likely to gain in importance in the modern world of digital information flows.
- Scholars or other professionals may meet at supranational conferences and explain their ideas to each other. All members of this world-wide “technological community” then tell their colleagues back home what they have learnt out there.

Of course, other variants may be thought up, and the ones listed may be further subdivided according to whatever criteria. At this point, however, there is no need for completeness. It is rather more important to realise that there are numerous channels through which technical (and other) knowledge can flow among countries and regions. We shall return to this matter at many points of the lecture.

1.2 Consequences of Globalization

1.2.1 Economic Growth

Trade in goods and services, the mobility of labour and capital, the diffusion of technical knowledge – all can and are likely to affect the growth performance of the economies that are involved. Assuming that the output of an economy can be represented by an extremely simple and stylised production function of the type $Y = Y(L, K, T)$, where Y ("yield") is defined as national output, L as labour input, K as capital input and T as an index for the state of technical knowledge, it is clear that the various types of flows across borders do have an obvious relevance for the economy's performance and growth. Hence Globalization matters: While the mobility of labour and capital change the amount of inputs directly, the state of technical knowledge and specialisation through trade do affect the degree of efficiency to which inputs can be transformed into output.

Of course, the statement that Globalization matters for economic performance and growth is still a rather trivial one. Just as the writing down of a production function, it is more an exercise in structuring our thoughts than in delivering a satisfactory explanation. The reason is simple: Trade in goods and services, the mobility of labour and capital, the diffusion of technical knowledge – all channels of Globalization have themselves deeper causes, which have to be identified. In econometric language: they are endogenous, and we have to look for exogenous variables to really come up with a meaningful interpretation of the facts. It will be one of our main tasks to carefully search for "deep causes" that together can serve as a satisfactory explanation or at least a meaningful interpretation of what we observe in the historical record and in present times.

A simple historical example may clarify what is at stake. One of the first and most fascinating cases of the link between trade and growth is Venice, the greatest Mediterranean maritime power of the 12th to 15th century. Venice was a trading city: it imported spices, silk and porcelain from the East (Byzantium, China, India) and grain from the West (Sicily); and it exported

– apart from the salt of the lagoons – manufacturing goods like glassware and textiles (based on raw silk) as well as ships and boats that were built in its famous shipyard, the “Arsenal”, the first and only large-scale industrial plant of the Middle Ages. Parallel to its increasing trade integration, the city experienced a sustained economic growth. People from its northern Italian hinterland “immigrated” to Venice, swelling its population size and making it, with more than 100,000 inhabitants, one of the largest cities in the contemporaneous world. The manufacturing of products with a high value added increased sharply, and so did the level of skills of the specialised craftsmen who worked for the export industry, be it in glass or textile production or in the shipyard. Naturally, much capital was invested and a rudimentary banking system developed to finance the trade ventures. Innovations in technology and organisation proliferated, and so did labour productivity all round – making the city one of the richest in the world.

In short: Venice delivers a wonderful story of what may be called a “virtuous circle” of economic growth, in which trade on a global scale, trade-induced specialisation of production and skills, immigration of labour, a nascent market for banking services and technical innovations nicely come together. It will be our task in the lecture to disentangle the causal chain of growth to find out what are likely to be the causes and the consequences of global integration. Of course, the rise and decline of Venice will be just one example on our agenda, though maybe one of the most colourful ones.

1.2.2 Cyclical Fluctuations

Economic growth is not a smooth process. It rather proceeds in cycles, which have many different causes and characteristics – from simple business fluctuations due to changes in aggregate demand up to waves of technical innovations that lead to an uneven time profile of productivity changes and investments.

To some extent, these cycles can be viewed as something natural, which does not really require a big explanatory effort, just like, e. g., the fact that human beings – in their time of adolescence - do not grow smoothly, but in waves and spurts. However, this calm and sanguine attitude is certainly less convincing for more violent fluctuations of economic growth. These do call for an explanation. Powerful booms and sustained recessions, and the path the economy takes towards and away from them, are very important phenomena with wide-ranging consequences that go well beyond the mere economic sphere. E. g., the Great Depression in the early 1930s was a major cause of the rise of National Socialism in Germany with all further disastrous social and humanitarian consequences.

This lecture is mainly on Globalization. Therefore, we will focus on cyclical fluctuations only to the extent that they may be caused, accentuated or attenuated by forces that are linked to Globalization. An obvious, but not the only candidate is the international mobility of financial capital. It is, e. g., broadly accepted that the Great Depression in the early 1930s began in the United States and was transmitted across the Atlantic via short-term international capital movements that led to a credit crunch in Europe and notably in Germany, which mostly depended on American capital. At that time, the financial crisis had disastrous consequences for the real economy in terms of a deep recession and sharp rise in unemployment. Most recently, in 2007/8, the so-called subprime crisis in the United States has also major repercussions in European financial markets because this time, European banks made heavy losses with American assets. Whether this will spill over into the real part of the economy remains to be seen.

1.2.3 Income Distribution

In the course of economic growth, there may be winners and losers, be it in absolute or in relative terms. And the fact that there are winners and losers may be linked to the forces of Globalization. Clearly, due to their economic and political importance, the questions of who are the winners and who the losers and why this is so and what could be done about it will stand at the centre of this lecture. We shall return to them on numerous occasions.

There are many channels through which Globalization may affect the distribution of national income. Specialisation and structural change that follow from trade may work in favour of skilled and to the detriment of unskilled workers, as is suspected in the present debate on labour markets in highly industrialised countries like Germany. An international capital market may be wide open to the well-informed smart professional, but much less so to the humble small-scale saver. Thus, in times of high inflation, e. g., the professional may avoid the debauching of his savings by shifting his portfolio into a foreign currency while the “ordinary small saver” may fall victim to a drastic devaluation of his modest wealth as famously happened in the times of German hyperinflation in 1922/3, with dramatic consequences for German society and politics.

Between countries, it is all too obvious that the fruits of Globalization are very unequally distributed over the world. E. g., there is a broad consensus that in the last two decades, China and India and other Asian countries have much more profited from Globalization than African

countries, which fell back in relative, some even in absolute terms. The reasons why this has been so will be discussed in Part V of the lecture.

2 Early Trends of Integration

In Part 2 of this manuscript, we take a look far back in economic history. We analyze early trends of integration through trade, labour and capital mobility and knowledge diffusion. In this context, “early” means: pre-industrial, i. e., the period from the beginning of human economic history up to roughly 1820, which is about the time after the Napoleonic wars when the so-called industrial revolution gained momentum, and so did long-distance trade in industrial products.

Note at the outset that, in this part of the lecture, we tend to use the term “integration” instead of “Globalization” because what happened until the 19th century appears to be much more modest than what happened later. It covered not really the world in any meaningful sense, but parts of it, though always quite prominent parts that play a major role in the history books. As we shall see, however, the partial geographical character of integration should not distract us from very important common elements to present-day developments. In this sense, we do really look at forerunners of Globalization that are important to understand the modern world.

Needless to say that the history of mankind up to 1820 is a fairly long span of time, and it may sound preposterous to cover all these vast millennia. But we have one excuse: the narrowness of our subject. Focusing on “integration”, we do delimit the subject in a reasonable way, provided that we take “trade” to refer to greater geographical distances, say, hundreds or thousands of kilometres, i.e. the sort of distances that roughly correspond to our modern notion of “international” trade. The same holds, mutatis mutandis, for capital and labour mobility as well as the diffusion of knowledge. With this limitation in mind, we can safely start with the Phoenicians and the Greeks, i.e. no earlier than the late second and the first millennium BC.

To be sure, beginning not earlier than the late second millennium BC does not mean that, before that time, nothing important happened in economic history. In fact, major structural changes had already taken place before, though at such an extremely slow pace that it defies the term “growth” in the modern sense of the term. After the end of the last ice age (about 10-12.000 BC) mankind had moved from a Paleolithic state of nomadic hunting with most primitive stone tools and weapons to the Neolithic state of sedentary agriculture, culti-

vating plants like wheat, barley, lentils and peas and domesticating animals like sheep, goats and cattle. Gradually, from about 6000 BC, a rural village life emerged, technology improved and new crafts such as baking, pottery, and even metallurgy began to develop. In general, first traces of local specialisation became visible, and so did a still very modest degree of urbanisation.

From the third millennium BC, one can speak of genuine urbanisation in some parts of the world, notably in Egypt along the Nile and Mesopotamia along the rivers Euphrates and Tigris. There the first empires emerge, and they share a remarkably high level of civilisation. Among many inventions and innovations, they created the first written forms of language: the Sumerian pictographs to keep a bureaucratic record of tributes and taxation. They made enormous progress in agricultural techniques as they developed sophisticated drainage and irrigation systems to maximize land yields from the fertile river marshes they lived in. And they developed a first sophisticated division of labour on the local level, with artisans specialising in textile and pottery manufacturing, metalworking and many other crafts.

What was largely lacking in these civilizations, however, is long-distance trade. Sure enough, there were occasional trading missions into far-away lands, but there was nothing like a well-organised trade network that might have led to a sustained interregional pattern of specialisation. It is not altogether clear why this so. One reason may lie in the fact that the various economies of Egypt and Mesopotamia were typically dominated by powerful state bureaucracies, which set other priorities than the typically decentralized commercial activities that underlie trade ventures. However, as we shall see in other parts of this lecture, even powerful centralised states do sometimes place their priority on trade if only because there is much profit in the form of taxes and tariffs to be extracted from it.

More importantly, the lack of long-distance trade may simply be the economic consequence of very high transport costs: potential trading partners in India or across the Eurasian landmass were few and very hard to reach by land or sea so that, given the technologies of the time, there was simply no economic basis for whatever form of long-distance trade integration. This did not change until the Mediterranean world appeared on centre stage of development and growth, and that was not really before the late second and the first millennium BC when the Phoenicians and then the Greeks moved into the foreground of economic history.

2.1 The Ancient World of Trade

2.1.1 Phoenicians and Greeks

The Phoenicians are the first people in the world whose economic history is strongly linked to trade and the specialisation that goes with it. Apparently, their geographical origin goes back to regions bordering the Persian Gulf and the Red Sea where they may have served early on as traders between Mesopotamia and Egypt. About 3000 BC, they moved to the Eastern coast of the Mediterranean, i. e. the region of present-day Lebanon, Israel and western Syria. They lived there in autonomous city-states, the most famous ones being Sidon and Tyre. As specialised merchants and sailors, they took advantage of the relatively low costs of maritime transport, thereby taking over the trade of the Egyptian pharaohs, whom they served as agents and contract merchants. As professional traders, they made major inventions and innovations, the most notable being the alphabet, which was more efficient to use in merchant business than earlier forms of writing.

From about 1200 BC, the Phoenicians succeeded in establishing a dense network of trading links all across the Mediterranean, thus becoming the dominant trading power of the contemporaneous world. As to the trade structure, the most renowned export items of the Phoenicians were transparent glass, of which they are supposed to be the inventors, high-quality timber from Lebanese cedars, which was widely needed for shipbuilding, and a famous purple powder, which was made from sea-snail shells and used by wealthy Greeks to dye clothes and garments. In fact, in Greek, the word “Phoenicia” means just land of the purple dye. On the import side, Phoenician trade consisted mostly of grain and foodstuffs as well as raw materials – silver from Spain, tin from Britain and copper from Cyprus, which was needed to make a durable metal alloy bronze. Given this structure of trade, the Phoenicians developed the nucleus of a manufacturing industry in glass-making and metallurgy, which was highly advanced and productive by the standards of the time.

Population pressure and the need for trading posts made the Phoenicians found many colonial cities, notably in the Western Mediterranean (see FIGURE 1). Among them are cities, which survived until today, e. g. Cadiz, Cartagena and Malaga in today's Spain, Tangier in today's Morocco and Palma de Mallorca, Cagliari on Sardinia and Palermo on Sicily. Many of these colonies became independent city states such as, most prominently, Carthage in today's Tunisia.

All these colonies – just as the “mother country” of Phoenicia proper – did not stretch far inland from the coast so that, by and large, the respective Phoenician hinterlands all around

the Mediterranean remained in an economically backward Neolithic state, well below the living standard and technological level that was attained in the coastal cities. This kept the military and fiscal costs of colonization low, but it made the Phoenician presence rather vulnerable. Once there was a strong enemy with a powerful land-based army, the strategic weakness was to become apparent. This is likely to be one of the major reasons why, eventually, the Phoenicians had to succumb to the military might of major land-based powers, the Macedonians under Alexander the Great in the fourth century BC and Rome in the course of three Punic Wars, which ended with the complete destruction of Carthage in 146 BC.

The Greeks are the second people of ancient times whose economic history is strongly focused on trade. Although the Greeks started as cultivators on the narrow stretches of arable land in their mountainous homeland, they soon began to engage in maritime trade, thus taking full advantage of the many natural harbours and islands of their “home waters”, the Aegean Sea. As merchants, the Greeks developed in a similar way as the Phoenicians, but they reached their peak of expansion a bit later (after about 800 BC). Firstly, the Greeks succeeded in dominating the trade in the Aegean and the Black Sea. Secondly, they founded a large number of colonies to further their trading interests and to alleviate population pressures that were felt most acutely on the Greek rugged peninsula where fertile land was extremely scarce. Virtually all Greek colonies were commercial centres and ports at the coast of the Mediterranean (plus, for that matter, the adjacent Black Sea). The philosopher Plato was right when, in the fourth century BC, he stated bluntly: we Greeks are sitting around the Mediterranean like frogs around a pond.

In its importance for Western civilisation, the spreading of Greek colonial cities around the Mediterranean can hardly be overrated. From the West coast of today's Spain to the East coast of the Black Sea, numerous Greek cities developed, usually sponsored by one mother city-state in Greece. As a casual glance at a historical map shows, the greatest density of Greek colonisation was achieved in Southern Italy and Sicily, around the Western and Southern coast of the Black Sea, the French Cote d'Azur and today's Libya (FIGURE 1). Among the colonies are cities that are still today big centres of trade and commerce, e. g. Marseilles and Nice in France, Naples, Syracuse and Taranto in Italy, Istanbul (Byzantium) in Turkey, Odessa (Olbia) in Ukraine and Alexandria (Naucratis) in Egypt. In fact, the dense network of Greek colonies delivered the basis for the cultural penetration of Hellenism that was so important for European civilization after the Macedonian empire had destroyed the powers of Athens and other Greek city states in the fourth century BC.

As in the case of the Phoenicians, maritime trade had a major impact on the pattern of specialisation of the economies of mainland Greek city states, at least of those like Athens with its port Piraeus and Corinth which were open to the sea. Grain cultivation was replaced by horticulture of grapes, olives and wine, which were better suited to mainland Greek soil and climatic conditions. They delivered a high value added per unit of input and per unit of weight and thus could profitably be exported to regions around the Mediterranean. In turn, grain and other bulky agricultural products were imported, along with raw materials as inputs into various crafts like pottery and metalworking that formed the basis for a growing manufacturing sector which, by the standards of the time, operated with highly skilled labour and modern techniques.

Complementary to trade, commerce, crafts and manufacturing, some rudimentary forms of financial services emerged in the major Greek cities, notably in Athens at about the time of its most forceful economic and political development, i. e. in the fifth and fourth century BC. Also, the use of coined money became much more widespread than it used to be. In fact, the monetisation of the economy was greatly enhanced by the recourse to the rich silver mines at Laurium on the Attic peninsula. The mines, which were state-owned by Athens, did allow the city not only to run substantial deficits in its trade balance over a considerable period of time, but also to finance lush public buildings and monuments, which – even as ruins – are nowadays major tourist attractions.

Politically and militarily, the Greek city states and colonies suffered from similar weaknesses as the Phoenician ones. As maritime powers without an extensive hinterland, they were eventually conquered by enemies who had stronger armies, but stood on a lower level of civilization and had almost certainly a lower per-capita income, if that had been measurable at the time. This was even more evident for the Greek than for the Phoenician case: when, with the battle of Chaironaia in 338 BC, the Macedonian king Philipp II set an end to the proud Athenian democracy and independence, many contemporary politicians – notably Athen's greatest orator Demosthenes – knew exactly what was at stake. But his desperate, almost life-long political attempt to put together a strong coalition of resistance against the Macedonian aggressor finally failed, not least because the Greek city states were not ready to put aside parochial interests to defend their political freedom. On the other hand, it is not clear whether political independence – if successfully defended – would have substantially altered the economic prospects of the Greek city states. E. g., Athens (with its port Piraeus) continued to play the role of a major commercial centre even in Hellenistic times and also, at least for a while, as a part of the Roman empire.

Be that as it may, there is hardly any doubt that trade integration and concomitant specialisation did help the two great maritime people in the Mediterranean – the Phoenicians and the Greeks – to achieve a relatively high level of prosperity. Whether their trading partners gained or suffered is very hard to say, given the scarce economic evidence on the matter. The colonies at least appear to have done fairly well.

There are even harder questions to answer, if we think of Phoenician or Greek society as a whole. Did the benefits of trade trickle down from the rich to the poor strata of society? Did many gain or just a thin elite of traders and craftsmen? To answer these difficult questions would go far beyond the scope of our lecture. After all, in ancient societies, work was done not only by free workers, but also by slaves, which enormously complicates any statement about the level of prosperity and its distribution. In the following sections on the Roman empire and its decline, we shall return to this matter in somewhat greater detail.

2.1.2 Roman Empire

It is necessary to put a note on terminology at the beginning of this section. For our purposes of analysis, which are mostly economic, it is convenient to speak of the Roman empire as the territory that was governed, administered and militarily controlled by Rome in classical antiquity. Hence our usage of the term is much broader than the standard political definition, which takes the empire to begin with the first genuine Roman emperor Augustus (27 BC-14 AD), i. e. after the demise of the Roman Republic, the military dictatorship of Julius Caesar and the transitional period that followed Caesar's death in 44 BC. Incidentally, our usage follows the German tradition of distinguishing between the "Reich" (in Latin: the imperium) and the "Kaisertum", which denotes the autocratic political system that followed the (aristocratic) Republic. Unfortunately, this distinction cannot sensibly be made in English because translating "Reich" into English yields "empire", which by definition is governed by an emperor and not by an aristocracy.

So much for terminology of the Roman empire and back to economics. There is a general consensus that, in the ancient world, it was the Roman empire that reached the highest degree of trade integration and, if we like, "Globalization". The peak was attained in the first two centuries AD, at some time between the reigns of the emperors Augustus and Marc Aurelius. And it was a peak that was not reached again thereafter for more than a millennium.

Remarkably, it was also a long-term peak in the standard of living. But was it really a high standard in any meaningful sense? We are tempted to answer yes when inspecting the im-

pressive remainders of Pompeii and Herculaneum, the two cities at the Gulf of Naples that, in 79 AD, were buried under the dust and lava streams of the volcano Vesuvius. Still today, this rich archaeological heritage gives a strong intuition for the high level of urban civilisation that was reached at that time in two small, though admittedly quite wealthy cities of the empire.

More reliable statistical evidence of Roman living standards is, of course, much harder to come by. A courageous attempt at quantification has been made most recently by the late Angus Maddison, a renowned economist and economic historian, who was specialised in the calculation and comparison of national income data over very long periods of time. On the basis of scattered evidence on many diverse economic indicators, he carefully estimated per capita income of the Roman empire at the time of emperor Augustus (actually 14 AD). He then compared his estimate with a well-known calculation for England and Wales in 1688, the earliest that is available for any part of Europe in modern, though pre-industrial times. His conclusion is that per capita income in the Rome of Augustus was about 40 per cent of the English level in 1688.

Is that much or little? It just depends on how you look upon it. Clearly, it shows that the true average of the population of the Roman empire was still significantly poorer than the population of one of the most advanced European countries even before that very country “took off” economically to become a pioneer of the industrial revolution in the 18th and 19th century. Other social statistics seem to support this view: e. g., life expectancy was not more than 25 years in Roman times, considerably lower than in England around 1700 where it stood at about 38 years (though these numbers raise their own problems of interpretation as they are, e. g., extremely sensitive to the social groups considered in the sample).

On the other hand, comparing the “true average” living standard in the way Angus Maddison does may seriously distort the picture if we want to know how a free Roman citizen of the first century AD lived relative to a pre-industrial average European citizen in the late 17th century. This is so for at least three reasons.

Firstly, at the time of Augustus, about one fifth of the Roman labour force consisted of slaves. Most of these, though not all, were much worse off than freemen. Hence an extremely uneven distribution of income in Roman society may mask the fact that the typical free urban citizen of the Roman empire – the person historians have in mind when browsing through the remainders of Pompeii and Herculaneum – was much better off than the average suggests.

Secondly, the Italian part of the Roman empire was richer than the non-Italian parts, with the difference being substantial if we can rely on the sketchy evidence on interregional income gaps in the empire, which has also been put together and analyzed by Angus Maddison. Hence, speaking of the empire as a whole does not mean speaking of its more wealthy heartland, which we mostly have in mind.

Finally, the standard of comparison – England and Wales in 1688 – is a very ambitious one indeed. In the 17th century, England and Holland were by far the most wealthy parts of Europe, much richer than most of the rest of the continent. Reaching 40 per cent of their standard means reaching a much higher ratio to the European average in the pre-industrial 17th century.

In sum, there are lots of open questions left in these types of very long-term comparisons of living standards, even if they are carried out with the utmost scrutiny. Be that as it may, however, it appears that, at the time of Augustus, the people of the Roman empire – or at least the free citizens of its Italian heartland – were not all that far away from the average living standard reached in Europe about 17 (!) centuries later. Or to put it boldly: 17 centuries of history seem to have been largely lost for the great goal of mankind to improve living conditions! To be sure, such a bold claim must not be taken too literally, as we shall see in later parts of the lecture. After all, these 17 centuries were not a dull stagnation, but rather a gradual fall followed by a gradual rise in economic performance over many centuries so that the picture of an awful loss of time is somewhat inadequate. Still then, it does deliver a first indication of how advanced ancient Rome already was, and what might have been spared mankind if the Roman empire had survived and realised a maybe slow, but sustained economic growth over time. We shall come back to this point.

What was the basis of the empire's economic strength? As in the case of Greece, it was to a large extent the trade network around the Mediterranean, which allowed a highly efficient division of labour over long distances. Other than the Phoenicians and to some extent the Greeks, the Romans were not a maritime people from the beginning, and long-distance trade never became a core matter of their identity. However, due to their successive conquests of practically the whole Mediterranean region as well as the Southern parts of today's France, Britain and Germany, the Roman empire governed precisely that share of the world where trade links had been much intensified over centuries. In fact, after the victory over Phoenician Carthage and the submission of the Greek peninsula as the Roman province Achaea, the Roman empire was the immediate heir of the prior maritime powers.

Typically, the rulers of the Roman empire kept the economic, social and cultural infrastructure of the conquered regions largely intact. More than that: it further developed old (and added new) maritime bases for trade in the Mediterranean region, including the adjacent Red and Black seas; and it made the sea routes much safer by successfully fighting piracy, the most ubiquitous threat to maritime trade over many centuries. In a sense, Rome became the caretaker of the Hellenistic Mediterranean world, rounded up by conquests in mainland Europe, which were called “Barbarian” but also treated with a certain respect for local customs and traditions. In return, the Barbarians – once members of the empire – did quite easily take over a somewhat Roman style of living, just as many immigrants into the United States in the 19th and early 20th century who were happy to become an ingredient of the American “melting pot”.

FIGURE 2 shows the Roman empire at the peak of its extension under emperor Trajan (114 AD). Juxtaposing FIGURE 1 and 2, it is fairly obvious that the Roman empire united virtually all relevant trading cities of the Mediterranean world under one administrative roof. While, legally, the status of the various conquered regions differed (from genuine Roman land down to Roman provinces), the economic treatment of the trading posts was similar everywhere. Most locations profited strongly from what is called the “pax romana”, a time of political stability and peace, of relatively free trade, competent administration and a coherent system of “Roman” law, which deserves to be called “commerce-friendly”. In particular, it ensured freedom of private contracts and, once contracts were concluded, a quite strict law enforcement. Even by modern standards, Roman law and its application provided a remarkably reliable and sophisticated legal framework for economic interaction.

Of course, in all different aspects, the “pax romana” was far from perfect: depending on time and places, there were violent local wars as well as poor law enforcement and rampant corruption, especially in the first century BC during the lifetime of Caesar, Cicero and Pompeius, when the Republic approached its end in massive political turmoil. Compared to other historical periods and places, however, the degree of military, political and administrative stability was remarkable indeed. And that helped to turn the whole Roman empire – centred around the Mediterranean – into an economic space that offered ample scope for a fully-fledged trade integration with all its productivity-enhancing consequences. This was most notably so in the first two centuries AD, from emperor Augustus to Mark Aurelius, but – with qualifications – it also applied to a much longer span of time, say, the two centuries before and after that peak period of the “pax romana”.

On top of these favourable institutional conditions, the Romans undertook massive investments in infrastructure, which also helped to facilitate trade and commerce. The most outstanding case in point is the network of long-distance land roads that were made of stone – the “Via Appia” being the most famous one that partly survived until today. To be sure, these roads were built for military purposes, which explains why they turned out too narrow for large commercially used carts. For light merchandise, however, they did reduce the cost of land trade, which in these days was very high indeed. As a result, the major Mediterranean shipping lanes were at least complemented by solid land routes, which linked more remote places to long-distance trade. Naturally, these land routes never attained the economic importance of the sea links. Nevertheless, they provided a first robust non-maritime infrastructure, something that had not existed under the Phoenicians and Greeks with their exclusive focus on the sea.

What was traded? Not surprisingly, the pattern of long-distance trade within the Roman empire largely resembled the structure that prevailed in the heyday of the Greek and Phoenician city states. Agricultural products dominated, with all coastal regions of the Western and Eastern Mediterranean exporting wine, oil, olives and related horticultural products and importing cereals from North Africa, Egypt and above all Sicily, which was then a major granary that supplied mainland Italy including the city of Rome with those large quantities of wheat that were badly needed to feed the population on a mostly vegetarian diet. Typically, upon arrival, the grain was stored in huge halls, which have partly survived as the ones of Ostia, the port of Rome. In many cities of the empire, crafts and manufacturing developed, and so did the export of cloths and textiles as well as ceramics and pottery, with the renowned “terra sigillata” from today’s city of Arezzo being a prominent example.

There was long-distance trade not only within the empire, but also with the rest of the world. Trade with the “Barbarian” North followed a pattern that, in qualitative terms, will re-emerge in the Middle Ages: textiles, ceramics, glassware, weapons and other relatively sophisticated manufactures going from South to North, and amber, grain, hides and furs, i. e. mostly products from agriculture or hunting and raw materials heading from North to South. Besides, the Roman empire imported Germanic slaves, which were traded in many Roman towns North of the Alps and then used either in Mediterranean agriculture and mining or in various urban crafts and services.

Trade with the East – Arabia, India and China – was different, at least on the import side. It amounted to the import of a large array of luxuries that were highly esteemed by the urban elite of the Roman empire: spices, incense, ivory, and other high-value items were ex-

changed for the standard Roman horticultural and manufacturing products. Note that, of course, long-distance trade way beyond the limits of the Roman empire was very risky and required extremely long travel times. E. g., the maritime round trip via the Red sea to some western Indian port to contact traders from China took a full year because one had to wait for appropriate north-eastern winds in autumn and south-western winds in spring to carry the ships (all, of course, sailing boats) forth and back.

Just as in earlier Greek times, long-distance trade incited a remarkably fine division of labour to develop. Specialized crafts and manufacturing flourished in virtually all cities of the Roman empire, but particularly so on the Italian peninsula. Short-distance trade – mostly on land routes – supported this trend. As cities grew into ever larger dimensions, there was an increasing need to supply them with fresh food (notably vegetables and fruits) from the neighbouring agricultural areas so that commerce, crafts and markets received an additional growth push. This, in turn, further supported urbanisation and the concomitant division of labour, with all standard gains from trade and specialisation to be reaped.

In fact, Roman society in imperial times became highly urbanized, at least by historical standards. Recent calculations indicate that modern pre-industrial Europe reached the degree of urbanisation of Augustus' Roman empire not before 1700. For Rome itself, historians estimate a population of between 350.000 and one million at that time, which makes it clearly the largest city of the world until Paris and London grew into similar dimensions in the course of the 17th century. Even the second largest city of the empire, Alexandria in Egypt, whose size is estimated at about 200.000 people in Augustus' times, would have remained at this historical place until the 16th century, when Naples, Paris and London finally overtook it.

In view of these facts, it is safe to conclude that the Roman empire had attained a degree of urbanisation that, in a much later period of history, was characteristic of a pre-industrial stage. Note, however, that this later period led straight into an unprecedented growth surge, which dramatically changed the world until today. After all, we must not forget that the span of time from 1700 to today is just three hundred years, roughly as long as the one that separates Augustus from Constantine who achieved a last consolidation of the empire. His was a period of history when the Roman empire had a standard of living that was certainly not higher than at the time of Augustus. If anything, it was lower because a century of big military threats and disturbances had past and left a very high tax burden on the population.

Note that, up to the economic peak of Roman imperial times in the 1st and 2nd century AD, the process of urbanisation and specialisation still proceeded – and so did economic growth,

at least as far as we can verify from the sketchy and mostly qualitative evidence that is available. Of course, by today's standards, it was very slow growth indeed, but growth it was. Remarkably, it was also a time of a moderate population increase, from below 30 million people in 300 BC to about 40-45 million people in the two centuries AD (estimated by Maddison and others for the territory of the Roman empire in the boundaries reached by Augustus in 14 AD). Apparently, for a historical while at least, there was some moderate economic dynamics, which also allowed population to rise, but that did not last. Bluntly speaking: The start was good, but the race was not continued.

Clearly then, a major question arises: why did the Roman empire not embark on a sustained growth path only remotely comparable to the economic upsurge that took place at the dawn of the modern industrial era? Apparently, major assets were present in imperial times that are today considered as to be most conducive to growth: a high degree of urbanisation and a fine division of labour in a well-organised and stable trading environment, based on a remarkably advanced physical infrastructure and a commerce-conducive legal framework. Even a rudimentary banking system based on a widely accepted currency existed: bankers ("argentarii") took deposits and provided credits at interest rates that varied widely, but were usually in the range of 6 to 10 % per year. Note that credit contracts entailed the enslavement of the debtor in case of his insolvency which – though reprehensible from a modern moral standpoint – was a clear indication of a very rational economic attitude vis-à-vis all forms of credit, be it for consumption and commercial purposes.

The puzzle becomes ever more striking when one takes the Roman level of technology into account. There is no doubt that this level was quite advanced by any reasonable historical standard. And it had made quite some progress over time. No doubt, the Romans had become masters in building and construction as can easily be recognized even today from the archaeological remainders of their cities. Their achievements in constructing infrastructures for urban and communal life as well as long-distance traffic are legendary and have partly remained tourist attractions until today: e. g., a permanent fresh water supply being carried by pipelines and over impressive aqueducts to the cities, elaborate systems of heating for water and buildings, roads made of solid stones over thousands of kilometres as well as innovations in shipbuilding and seaport facilities. Also, in crafts and manufacturing, techniques were improved in virtually all relevant fields – from metalworking to the making of glass, ceramics, pottery and textiles. In the urban workshops of specialised weavers, e. g., a vertical loom was used since Greek times and later technically upgraded by the Romans.

Given this advanced state of technical knowledge, why then did not a further dynamics evolve that might have led to sustained growth and path-breaking innovations with productivity-enhancing effects? Of course, to answer to this question is largely a matter of historical speculation so that any attempt must be treated with due caution. Be that as it may, economic historians have tried to give an answer, and we try to do so as well, thereby roughly following the mainstream line of reasoning in the literature.

The key to the puzzle appears to lie in agriculture. Despite the high degree of urbanisation, agriculture was and remained by far the most important economic activity in the Roman empire, both in terms of output and employment. This itself was a kind of indicator for the lack of sustained growth because, to go well further beyond the (admittedly high) degree of specialisation reached by Rome, a still higher surplus would have to be extracted from agriculture than was actually done. Clearly, the growth pushes in later periods of history, notably the times of the Middle Ages from the 11th to the 14th century and the so-called industrial revolution in the 18th and 19th centuries, had strong roots in agriculture. They were only possible because, due to major innovations (technical or otherwise), there was a sustained increase in agricultural labour productivity, which allowed to produce with less manpower more or at least an equal amount of food, the most basic need of mankind. Only that gave the leeway to massively release workers into other economic activities, which then quickly realised their own productivity gains through specialisation and innovations. This, in turn, fed back through many different channels into agriculture. We come back to this kind of “virtuous circle” a couple of times again in this lecture.

In imperial Rome, nothing like that happened. To be sure, major labour-saving technologies were known and used in mining, energy production and agriculture. And they were also seriously analyzed by great minds of the time, like Vitruv in his “de architectura” and Pliny the Elder in his “naturalis historia”. Most notably, the classical Greek plough was used and improved in Roman times. For its application in the more Northern provinces with their heavier soils, it was upgraded and equipped with wheels so that two oxen could pull it. Also, the water mill was well-known and applied, although a widespread use of it was limited to those areas where there was ample supply of water most of the year round, which was not the case in the many dry regions of the Mediterranean.

Clearly, there may be other examples for techniques that enhanced productivity and saved labour. Whatever the state of technology that was reached and applied in imperial times, however, there were apparently no further big breakthroughs that could open the door to major improvements of labour productivity in agriculture. Why was this so? Two hypotheses are

plausible: lack of labour scarcity and lack of entrepreneurial spirit. We look at them separately.

For a long time, labour was not scarce in ancient Rome. Successful wars and conquests provided a seemingly never-ending stream of slaves that could be used to supplement domestic labour wherever scarcities emerged, from services, crafts and agricultural work in the fields down to the most dreadful drudgery that the ancient economies had to offer, the work in the mines, which was almost exclusively done by slaves. Of course, the ample supply of slaves had indirect effects on the domestic labour market as it depressed wages of the free labour force. Hence, for a long time, there was indeed very little economic incentive to search for labour-saving technologies. And it is for this long period until about the early 2nd century AD that the lack of labour scarcity has a high explanatory power to account for the stagnation of labour productivity.

Gradually, however, the state of the labour market did change quite substantially. Under the rule of emperor Trajan in the early 2nd century AD, the Roman empire reached its greatest expansion: no further conquests followed, and most of the wars were not successful. Hence the forced supply of slaves from afar dried up. True enough, slaves could still be bought in the free market, but – with new supplies being curtailed – prices for them began to increase as well. And so did, eventually, wages for free workers, with the inelastic supply of slaves being one, but not the only one reason for the price of labour to shoot up. Apparently, population growth came to a halt due to relatively low birth rates in the empire, both of the free and the slave part of the population, with the latter having typically below average rates of reproduction. Parallel to these trends, the enormous weight of administrative and military duties of the empire absorbed more and more labour, and that which remained in economically viable work was ever more heavily taxed – thus also putting an upward pressure on wages.

Clearly then, at least in the later stages of the empire, the abundance of labour was dissipating and gave way to a genuine scarcity that somehow should have spurned labour saving innovations. Note that, by later stages, we mean not just a few decades of the empire, but quite a long period of about three centuries. After all, historians date the death of the Roman empire as a reasonably identifiable political, administrative and economic unit certainly not before the second half of the 5th century AD so that there was still plenty of time left to react to economic incentives. But the existence of economic incentives is one thing, the receptiveness to them is another.

This is the point where the second hypothesis comes in: lack of entrepreneurial spirit. Historians agree in their judgement that, in the ancient world of Rome (and also of Greece), entrepreneurship in the economic sense was not highly esteemed. No doubt, the Roman ideal of a gentleman in society did not contain any major dose of success in business. In essence, the ideal was confined to political, military and cultural traits. A gentleman ought to be respected in some sort of public role, however defined in detail, but largely excluding private business. And he should be able to fill this role on a full-time basis without pursuing any business-type goals.

Of course, to act as a gentleman in such a public role was voluntary and honorary. It did not yield an income stream. To the contrary, it usually involved expenses, sometimes heavy expenses as Roman history shows. Therefore a comfortable economic basis was needed, and that, after all, had to be created by some sort of business, be it agriculture, crafts or commerce. But this business was viewed more like an unavoidable nuisance that should be ended as soon as the individual economic conditions allowed. The ideal was and remained the gentleman who owned enough land to retire from business and live off the returns of his land, with no further investment in time and effort in improving yields or engaging in whatever commercial activity. And if, as often happened, the returns of land did not cover the expenses for a while, then it was standard (and honourable!) practice for a gentleman to resort to a consumption credit (as we would call it today). In fact, it was much more common in imperial Rome to borrow money for consumption than for commercial purposes. In a way, the modern saying “time is money” was read by the Romans from back to front: “money is time”, i. e. once you have enough money, you use it up as time for more important purposes than pursuing money-making. And if you don’t have enough, then you borrow some.

This anti-commercial attitude is reflected in many documents that have survived until today. Most prominently, it can be found in the writings of Cicero who virtually ranked the various professions according to their social respectability. In his judgement, which was not at all unworldly (he was not only an intellectual, but a lawyer and politician as well!), commerce and trade were mean occupations, though long-distance trade was “not so very discreditable”, a wording that indicates how mean indeed short-distance trade and commerce must have been in Cicero’s mind. And he was not alone with this view.

In a world with this type of values prevailing, an all too openly profit-maximizing business attitude is likely to be viewed with great suspicion: a merchant who makes big profits from arbitrage trading, a banker who charges high interest rates, but also a landlord or a farmer who searches for labour-saving devices and invests respectively – all of them will find a so-

cial climate that is anything but conducive to business gains and growth. A famous and often-quoted anecdote from the 1st century AD on the emperor Vespasian – whether true or not – fits into this picture: confronted with the offer to use a machine to facilitate weight lifting in some public construction works, he turned the offer down by saying that he would rather want to give work and food to the people. It is likely that this very climate dominated all throughout the Roman empire and set a narrow limit on its growth potential once the challenges of structural change in a world with increasingly scarce labour became acute.

Of course, even this kind of “sociological” explanation leaves us back with many puzzles and questions. How could societies like ancient Greece or Rome be so highly innovative in fields like arts and philosophy (the Greeks) and law and administration (the Romans) and nevertheless lack the acumen to adjust to new economic circumstances? Was that really so difficult? Why was it impossible to turn the scientific and technological curiosity, which was in ample supply in many Greek and Roman minds, into an economically viable tide of innovations? Why did the apparent success of trade, commerce and manufacturing not gradually shift the attitude of the leading Roman classes towards more respect for the commercial activities of merchants, craftsmen, farmers or whoever did business? Why did none of the great Roman minds develop and elaborate the idea (as Adam Smith did in the 18th century) that, after all, the pursuit of individual profit might be in the public interest?

All these questions must remain unanswered. However, they should remind us that the fading away and the eventual demise of the Roman empire as the first large-scale effort of economic integration (or, if we like, “Globalization”) was not an unfortunate accident of history. Rather it came about because a quite successful and open society did not react dynamically to the challenges of time. Trade, commerce and manufacturing flourished, but there was no internal mechanism in society to heave the economy onto a sustained growth path. And all this happened in conditions and circumstances that look quite favourable even from the hindsight perspective of today.

2.1.3 Sliding into a Dark Age

The decay of the Roman empire is a long and complex story. It is certainly interesting for the historian to analyze the various stages of agony and defensive struggle that the empire went through. In fact, the process of shrinkage has quite fascinating aspects: many border wars with Germanic tribes, a powerful de-urbanisation and, periodically, some courageous political attempts at consolidation. All this is not part of this lecture on Globalization.

It is equally tempting to speculate about which of the many particular hypotheses about the causes of the empire's decay is the most successful in explaining the actual facts. All this is also not part of this lecture. However, some brief notes are justified in view of our prior discussion in section 2.1.2 of this manuscript.

With some sweeping simplification, we may distinguish two types of explanations for the decline of the Roman empire:

- demographic causes, i. e. a shrinking population due to a declining birth rate, plagues and famines and a worsening climate; and
- fiscal and military causes, an ever rising tax burden due to ever more forceful military threats from outside by Germanic tribes.

Both types of explanation are plausible and may together deliver a reasonable descriptive account of the events that ended in the collapse of the Western part of the empire and the survival of a smaller Eastern empire (Byzantium) with Constantinople as its capital. As an economist, however, one is left with the uneasy feeling that they do not really identify the real "exogenous" driving force of the process. What appears to be lacking is exactly what has been discussed in section 2.1.2 of this manuscript: the inaptitude of the Roman empire to generate economic growth in the sense of a sustained increase in labour productivity. To some extent, poor demographic trends and crushing fiscal burdens are themselves the result of a lack of economic growth. This is quite easy to see with some counterfactual speculation:

- Had the Roman empire been able to initiate, e. g., a kind of agricultural and commercial revolution with new techniques and a major rise of labour productivity, the negative economic effects of demographic trends might have been mitigated and the trend itself possibly been turned around. After all, a higher surplus might have incited higher birth rates and a revival of migration into the cities, thus preserving or even increasing the already high degree of urbanisation and specialisation that had been reached. In fact, this is exactly how things worked in later more successful periods of history when the demographic tide turned (e.g., notably after the Black Death in the mid 14th century).
- Had the Roman empire been able to generate only a moderate economic growth over the 300 years of its decay, the absolute burden of taxation to finance military protection would have been – as a share of output – much less crippling. There are many modern examples of countries that carry heavy fiscal burdens, but can do so because of their economic dynamics. E. g., in the second half of the 20th century, many European nations built up and financed large welfare states, while Israel and the United

States incurred vast military expenditures. They all could afford it as long as their economic growth performance was dynamic enough.

Of course, in the end, it is a moot question to ask what is more “exogenous”: the failure to respond to economic incentives and generate growth, or the actual challenges that brought the collapse to a stagnating system. After all, it is a matter of perspective: economists, who look at the counterfactual, may emphasize the former, and historians, who look at the factual, the latter.

Whatever its deep cause, the decay of the Roman empire (or more precisely: its western part) had vast consequences. For our purposes of purely economic analysis, three stand out:

- Firstly, the decay of the empire put an end – though a gradual one – to the ancient long-distance trade network that had been successively set up by the Phoenicians, the Greeks and the Romans. It was the first rudimentary “Globalization” that the world had seen. A huge and intricate division of labour broke down, and with it all the positive productivity effects that could be ascribed to it.
- Secondly, the decay of the empire sharply reduced the degree of urbanisation: cities shrank dramatically, and with them the division of labour in the cities and between the cities and their neighbouring countryside. Thus even short-distance trade contracted down to a very low volume – so low, in fact, that the degree of professional specialisation must have reached a long-term minimum.
- Thirdly, the decay of the empire led to a drastic structural change of agriculture. Due to labour scarcities and the rising insecurity caused by the breakdown of administrative and military protection, traditional small-scale farming gave way to cultivation in large estates (in Latin: latifundia). In effect, peasants voluntarily left their land to search for the shelter of great landowners, who provided security and organized work more efficiently than smallholders could do on their own.

Note that these developments were gradual, stretching over centuries, not decades or years. Eventually, they led towards a kind of odd convergence of living conditions in the territory of the formerly rich empire and in Barbarian lands. It was convergence in poverty at a time that historians aptly call the Dark Ages. Not much is known from this period of history, but what is known indicates that living standards in the territory of the former empire must have been much lower than before.

Note also that it is the time of a gradual rise of feudal institutions: once the legal, administrative and military power of the empire faded, the economic demand for security was met by quasi-feudal lords who provided protection, but demanded labour services or some sort of taxes in return. This trend had a remarkable side effect: it undermined the institution of slavery. Without strict law enforcement and with labour in high demand, it became increasingly difficult to keep slaves against their will and to treat them all too badly. After all, the risks for a slave of fleeing from his master were drastically reduced and the likelihood of being accepted as a worker (or peasant) by a new employer (or landowner) much increased once labour became really scarce. In fact, there were attempts made to reduce mobility by forbidding slaves and peasants to leave their workplaces in a large estate, but they were hardly successful. Hence, ironically, the decay of the Roman empire had a kind of positive humanitarian side effect: the erosion of slavery. However, the new forms of labour organisation that followed with the rise of feudalism were not much more attractive. After all, slavery was replaced by serfdom, and that is still a long way from freedom.

2.2 Middle Ages and Renaissance

2.2.1 Early Feudalism

Historians put the end of the Roman empire into the second half of the 5th century. At that time, Rome lost the military, political and administrative control over the western part of its empire. Economically, things developed much more gradually, with the process of population decline and trade disintegration stretching over centuries rather than decades. However, some time between the 5th and the 10th century, a low point of population size and trade integration was reached in the western territory of the Roman empire. It was the nadir of the so-called Dark Ages, the time of early feudalism.

Very little is known and documented from the economic life of early feudal times. This by itself is an indication that some sort of subsistence agriculture must have dominated economic life. After all, it is the cultivation of land that leaves the least traces in records that could be studied by latter-day historians. In fact, the scarce evidence that is available does point towards an agriculture with a very low productivity. E. g., for the Carolingian period, i. e. the 9th century, which was certainly not the darkest of the Dark Ages, the average yield of harvests was in the range of one-and-a-half up to two times the seed. In other words: about half to two thirds of a harvest was necessary to just keep the status quo level of bread grain cultivation

constant, and only the rest – one third to one half – was available for consumption. By all means, this was certainly close to the bare minimum for survival in normal years and a ticket for starvation whenever a series of bad harvests came about.

How does it compare with the yields in other periods of pre-modern times? The answer is: miserably. E. g., in the Roman empire eight centuries earlier, the records speak of average yields of about four times the seed, with peaks of ten times on very good soils as those of Sicily. And for England in the 13th century, a yield of three to four times the seed has been calculated by historians of agriculture, which is of course still low by modern standards. Clearly then, in absolute terms and by all historical standards, agriculture in early feudal times was no more than subsistence agriculture. As such, it allowed in normal times – though not in bad years – the bare survival of people and eventually stopped the decline of population size at a low level. However, it hardly left any surplus that could be spared to support a significant non-agricultural expansion or even urban growth either by releasing labour that was not needed in agriculture or by providing savings to invest in other activities or in whatever urban infrastructure.

Trade, by all means, was an absolutely marginal affair. To be sure, many small villages, hamlets or farmsteads that lived from agriculture were occasionally visited by some merchants that travelled long distances to offer a collection of portable, usually high-value wares of which the most basically needed was probably salt for conservation purposes. The scale of this trade must have been very small indeed. Nevertheless, to keep it up in dark times was a major economic service to society as it provided a minimal long-distance network not only of commerce, but also of regular information. In a sense, the merchants were not only traders, but also news agents.

It is not surprising that, in these circumstances, merchants had a particular reputation of being somewhat shady and mysterious figures. After all, they came from far away, they carried exotic wares, and they told fantastic stories about foreign lands and alien people. Needless to say that, given the enormous risks of long-distance trade in this period, commerce was a most adventurous affair and thus attracted a type of people with exceptional courage, cunning and ruthlessness. Thus subjective myths mixed with objective facts and nurtured widespread prejudices about trade and merchants. Among these, Jews were prominent. And it is not farfetched to suspect that future anti-Semitism has one of its roots in this period.

Clearly, in this state of almost-autarky, the gap in living standards between the “Barbarian” north and the “post-Roman” south of Europe was bound to narrow. Nevertheless, it did not

disappear completely as the roots and remnants of urban life, where it had existed, were not altogether eradicated or destroyed. Also, some crafts tradition survived, and maybe some small-scale manufacturing, but most of it was of a coarse kind, directed at the most basic needs of country people for clothing, pottery, carpentry and the like. By and large, however, it was the productivity of a self-centred, almost autarkic agriculture that determined what – invariably low – standard of living prevailed and whether there was an opportunity to get out of this equilibrium of poverty at some date in the future.

It is one of the big puzzles of history that, in these Dark Ages, it was in fact manorial agriculture that became the seedbed of a first growth impulse. Apparently, things were not quite so static as they looked like. To be sure, it was a very modest dynamics - not yet comparable to the later Middle Ages, not to speak of modern times. If, e. g., we assume that, between 700 to 1200 AD, agricultural productivity doubled (which may not be far off the mark if we trust the estimates quoted above!), then this would amount to an average growth rate of about 0.14 % per year. This is not remotely comparable to the productivity dynamics we are used to nowadays, where any growth of whatever measure of economic performance at rates below 1 % per year is regarded as virtual stagnation. However, an average of 0.14 % productivity growth per year is not nothing, and sustained over a sufficiently long period of time, it can change society quite dramatically. In any case, it stands in stark contrast to the record of flat stagnation and eventually shrinkage of the mature Roman empire. Apparently, something was there in early feudalism that was not there in Roman times – something that gradually made the economy roll faster again and get onto a new road of integration and urbanisation.

What was it? To answer this question, we first have to briefly look at how the economics of agriculture in early feudalism worked – in terms of both institutions and technologies. Note at the outset that, in doing so, we exclusively focus on the economic side of feudalism. We leave aside all political, social and military aspects of feudal agricultural society as it paradigmatically evolved, e. g., in the Frankish Empire from the 8th century onward. In particular, we neglect the pyramid power structure from the king down to the landlords, which is at the core of the persistent power struggles of early medieval times that fill the standard history books.

For us, to understand the economic side of feudalism, it is sufficient to condense the political side in three major stylized facts that characterize the time: (i) an extremely low population density, (ii) a mighty and persistent military threat from outside by aggressive wandering people (e. g. Magyars, Muslims or Vikings) and (iii) a very high cost of control and information. All this pulled society in the direction of a loose kingdom with a kind of pyramid pow-

er structure where, unlike in Roman times, the king was unable to enforce a centralised taxation to raise funds for internal and external security, but had to resort to a system of providing the nobility with land that generated a sufficient yield to finance the individual landlord in his function as a mounted warrior. In essence, upper lords provided fiefs for lower lords ("vassals") in return for military services. For us, only the lower ends of the pyramid are important because it is there that, besides taxation and military service, the economic organisation and technology of agriculture comes in.

The core of the economics of feudalism is what historians call "manorialism" – derived from the word "manor", which goes back to the Latin term: "manere", i. e. to hold or to inhabit. A manor was a small agricultural village community, highly integrated and almost autarchic. Economically, it was a kind of optimal adaptation to an environment of desperate poverty, low population density as well as pervasive lawlessness and insecurity.

Essentially, the manor was based on an implicit contract between a landlord and "his" peasants: the landlord was to provide the public goods of security and law; in return, the peasant-serfs worked on his land (the "demesne" or, in modern spelling, domain) and/or handed over part of the produce from their many, but much smaller holdings, be it grain from cultivation or meat from livestock raising. In addition, there was some common land. Note that the peasants were serfs, not freemen: they were "attached" to the land and the village in the sense that they did not have the right to leave, but were also protected from being chased away by the landlord. Note also that, quite often, the landlord did not supervise the cultivation of his land himself, but left this to a bailiff or steward.

Clearly, the implicit contract underlying the economics of the manor was far from perfect in any sense, and one should not expect that anything remotely resembling modern levels of security and the rule of law came out of the arrangement, given the strong monopolistic position of a landlord as the only one heavily armed person in the village, and the chief judicial authority at that. Sure enough, the degree of "taxation" of serfs by the landlord also reached high levels: for England in the 13th century, it has been calculated that about 50 % of the peasants' income was transferred to the landlord in one way or another, and that is generally considered to be, if anything, a lower bound.

Nevertheless, in large parts of Europe, the implicit contract proved quite stable. Why was this so? Apparently, the public good of security had such a high survival value that, grudgingly, peasant-serfs paid their dues and levies and delivered their labour services. On the other hand, they were themselves not altogether without power. After all, labour was scarce and

thus had “hidden weapons” at hand that could be used to counter any all too extreme abuse of power by the landlord. As in modern times, various forms of collective action (strikes, passive resistance etc.) could be used to put pressure on the landlord, and so could the implicit threat of deserting the village, though this was illegal and far from easy under the conditions of the time. In normal circumstances, however, one should expect from peasants to have skilfully used the typical information asymmetries that underlie what we today would call a principal-agent-setting, with the lord as the “principal” and the peasant as the (better informed) “agent”. After all, it was not so easy for the lord to verify exactly what the harvest yields on the peasant’s land would be so that, to some extent, a peasant could avoid an all too heavy “taxation” by holding back some of the produce for private purposes.

Agricultural organisation on a medieval manor was quite different from ancient Roman time. In particular, the work was done not in enclosed, but in open fields. This was not only so on the common land, but also on the peasants’ holdings, which were arranged in long narrow strips that were not separated from each other. That was completely different from the Roman past of enclosed holdings that formed a checkerboard pattern. The open long narrow strips had a number of advantages in the prevailing circumstances: e. g., they provided a fair distribution with respect to the quality of land, and they allowed cattle to roam over a wider range once fields lay fallow – thus not only grazing and fattening, but also providing manure for the recovery of the soil.

At first sight, the work in open fields may not look like a very exciting break with the past. However, at least in the northern and central plains of Europe, it had important consequences for the productivity of land and labour once new technologies emerged. In fact, it laid the ground for massive economies of scale that could eventually be exploited. In contrast, the classical Mediterranean regions – with the important exception of the fertile Po-valley in northern Italy – did hardly profit from a shift to open fields. In fact, there were too many hills and mountains that prevented a strict ordering in equally long and narrow strips; and there was much less cattle that provided meat, milk and manure.

In view of these very different conditions, it is not surprising that, generally, the shape and structure of the manor had many local and regional variations. In its prototype form, as it is described here, it fits best into a western/northern/central European pattern, and less so into the Mediterranean and the Alpine regions. Remarkably, on marginal lands in geographically remote regions like the Alps, feudalism – and with it the dominance of the manor – never took strong roots, and a substantial free peasantry survived.

So much for the basic economics of manorialism. At least in the wide plains of western and central Europe including northern Italy, it proved remarkably stable. More than that, it proved to some extent innovative. During early feudal times, there were quite a few improvements in agricultural technique and organisation. Three major ones stand out: (i) three-course field cultivation, (ii) the use of horses as draught animals, and (iii) the heavy wheeled plough. They together led to a significant rise of productivity, at least in the longer run. We summarize them very briefly.

(i) **Three-course field cultivation** was a logical consequence of the conditions of climate and soil in the plains of western and central Europe. In the Mediterranean south (and in the Roman past), a field was cultivated and lay fallow every other year. This two-course rotation was well suited to the relatively dry climate and light soils of the Mediterranean, where moisture and fertility were the limiting factors that had to recover in a regular and relatively long time span of non-cultivation. In the more northern plains, with less scarcity of land and forests, the traditional form of agriculture first resorted to a primitive slash-and-burn technique, which still had a nomadic element: fertilizing the soil by burning the woods, then plant field crops until the soil is exhausted and then move on to do basically the same somewhere else. In Roman times, the two-course rotation was introduced, but the moister climate and heavier soils of central Europe allowed a more ambitious move to a three-course-system: sowing a crop in spring to be harvested in late summer (course 1), then sowing a winter crop in autumn to be harvested next summer (course 2), and then allowing a fallow period (course 3) to start all over again next spring.

To be sure, three-course field cultivation was not possible everywhere, and to develop its full potential it required improved techniques (to which we turn below). But where and when suitable conditions were met, it had obvious advantages: it was at least one third more productive than two-field cultivation, and it allowed different crops to be planted – usually barley, oats, beans and peas in spring, and bread grains like wheat or rye in autumn. This broadened the diet and provided fodder for horses and cattle. Most importantly, it reduced the likelihood of famines in case of a bad harvest; e. g., whenever the bread grain crop was a failure, a well-stored spring crop of last year provided a kind of emergency supply. Thus the three course rotation also figured as a most basic form of disaster insurance.

(ii) **The use of horses as draught animals** was unusual in ancient Rome and remained so in most of the Mediterranean region, which continued to resort to oxen for field work. The reason is simple: though horses are much stronger and faster than oxen, they are best fed with oats, which does not grow in most of the Mediterranean areas. This is why, in early me-

dieval times, a substitution of horses for oxen was only feasible in regions where oats could be harvested, i. e. in the western and central plains of Europe. Even there, however, a formidable problem had to be overcome: a horse was to be shod and harnessed in a way that it could in fact be used in heavy field work. In Roman times, horseshoes did not exist and the standard harness was such that it interfered with the animal's breathing (and thus strangled it when pulling weighty loaded carts or wheeled ploughs). To be sure, horses were then only used for military purposes so that no need for new technical solutions was felt.

This changed in the Middle Ages. Some time before the 10th century, two major innovations were made: the horseshoe to protect the delicate hooves and the horse-collar, which rested on the shoulder of the horse and thus allowed the animal to breathe freely, even when pulling heavy weights. Further improvements in the use of the horse in fieldwork followed, most notably the breeding of stronger animals, which had more power at the cost of being somewhat slower. In any case, vast new options for productivity gains were opened. It is remarkable, however, that even in the agriculture of western and central Europe, the horse did not completely replace the oxen, which is good proof of the tight cost and productivity balance between the two. It all depended on a variety of economic and environmental circumstances which of the two prevailed.

(iii) The **heavy wheeled plough** was introduced some time before the 10th century. It was not unknown in the Roman empire as a note by Pliny the Elder on cultivation methods in Gaul (today's France) indicates. However, it was probably not used on a broader scale at that time. After all, the light soil of the South could better be cultivated with a traditional Mediterranean plough, which just scratched the surface to allow for moisturising. In the heavy soils of the North, a plough that allowed to turn around the deep loam or clay was obviously superior, but it required several oxen or horses to pull it.

Sure enough, the heavy plough was something that, in today's language of economics, would have to be termed a key technical innovation, which had the potential of revolutionizing agriculture in terms of land and labour productivity. Note, however, that to unfold its full potential, major complementary innovations or at least adjustments were necessary, with the horseshoe and -collar as well as the three-course rotation being only the most prominent ones. Clearly, the peculiar forms of manorial organisation of agriculture were now also a great asset: the long narrow strips, which minimised the complex action of turning around a heavy wheeled plough with several animals; the cooperative work in open fields, which was quite suitable to the use of sophisticated tools like the wheeled plough; and the three-course

rotation, which allowed for a high degree of capacity utilisation of a capital stock that now also consisted of quite expensive animals.

All in all, it is plausible to assume that these innovations did not emerge and develop independently of each other. Given the paucity of the empirical evidence, we will probably never be able to disentangle the causal chain, which of them came when and in what shape and how it fed back into the design and use of the others by providing economic incentives for technical adjustments. Be that as it may, the totality of the innovations is the big achievement, and it raises a real big puzzle: how can a society so poor and isolated come up with a set of major organisational and technological improvements in agriculture?

As usual with big questions in history, the answer must remain speculative. However, as in the case of the stagnation of the Roman empire, two hypotheses come to mind: labour scarcity and mentality. As to the scarcity of labour, it is all too obvious that it was pervasive. Within any almost autarchic village, there must have been a strong incentive to do everything to minimize labour input. In this respect, maybe manorialism even had a distinctive advantage. In its cooperative effort, the village worked almost like a closed production unit, like a family-run medium-sized firm. Of course, that firm did not produce for the market, i. e. it did not sell its produce to make a profit or surplus. Rather its produce was consumed by its members (landlord and peasant-serfs) who, under the prevailing subsistence conditions, had an utmost interest in maximising the yield in order to maximise the likelihood of survival of themselves and their families. Hence, despite the virtual absence of commerce and trade, a strong business-like attitude re-emerged, and with it the incentive to introduce labour-saving techniques.

As in the case of ancient Rome, however, it is hard to dispense with a second hypothesis: mentality. After all, economic incentives are one thing, taking advantage of them another. In fact, the use of the heavy plough may be a good case in point: it was known (and occasionally applied) in Roman times, but the breakthrough came centuries later in Medieval Europe. Apparently, some receptiveness to technical change must have grown in between. Here, of course, the door for speculation is wide-open. Maybe that Christian values somehow shifted the general attitude away from political not only to spiritual, but also to economic goals and matters – in the sense of the monastic slogan “ora et labora”. Maybe that the lack of cities as the classical places for political discourse simply made the big lofty issues of the grand polity (the “res publica”) fade away to be replaced by a down-to-earth common concern for economic advancement of the small manorial village community. Or maybe simply that, once generations of people have been sufficiently poor and – again and again – lived through disastrous famines, plagues and a climate of shrinkage all round, they do quite naturally change

their values and priorities away from the “luxuries” of civilization towards the very basic improvements in living conditions. It all remains speculation, but a fascinating one at that.

2.2.2 The Great Expansion 1000-1350

Some time between the 8th and the 10th century, economic growth resumed in Europe. Historians tell us that, in this period, reports about the foundation of new villages and the clearing of forests became more frequent, which points towards an increase in population size and economic performance. Also, the climate began to warm so that new settlements were founded up to the northern rim of Europe. At the same time, two of the most fertile regions of Europe, the areas of Milan and Paris, showed the first signs of crowding. From the 10th century onwards, the evidence for growth is unmistakable. A sustained expansion set in and lasted until the first half of the 14th century when a series of plagues and diseases led to a sharp decline of population and a new period of labour scarcity.

As usual in the Middle Ages, dates of trend change cannot be identified precisely, but the time from the late 10th to the early 14th century may well be the first period in history that deserves the title “Great Expansion”, both in terms of demographics and economics. Of course, no reliable statistics are available to quantify the expansion; after all, the first population census took place in the city of Venice in 1338, and the first reasonable estimate of national income was the one for England in 1688. But historians are clever and dig deep to find evidence.

Drawing on a variety of unsystematic sources that have been carefully studied by medievalists – e. g., written records on inventories of manors, conscription lists for military service, various poll and hearth taxes as well as evidence from archaeology, botany and toponymy (i. e. the study of place-names!), – the Italian economic historian Carlo Cipolla claims that, for the territories of today’s Britain, France, Germany and Italy taken together, the population almost tripled in the period 1000 to 1300 – from 15 to 42 million, with the relative growth being strongest in the more northern and central regions, which had an extremely low population density to start from. For Europe as a whole (including Russia and the Balkans), Cipolla postulates a growth from 30-35 to 80 million people over the same period of time. All this amounts to an annual average growth rate of slightly above 0.3 %. Note that, from about 1300 to 1500, the same statistics show that population stagnated – as a consequence of a sharp decline in the 14th century and only thereafter a gradual recovery, which gained momentum not before the 17th and 18th century.

To sum up, there appears to be enough evidence for a radical demographic change between about 1000 and 1300, despite the paucity of the data. To be sure, an average growth rate of population of a bit more than 0.3 % per year is not much by the standards of the last two centuries. Historically, however, it was nothing less than a demographic take-off. Of course, even at the end of the expansion, population density was still extremely low in Europe – actually less than a fifth from what it is today! On the other hand, it was nowhere as low anymore as it had been in the Dark Ages.

Of course, a demographic take-off immediately raises the question why it happened. In purely arithmetic terms, there are two answers: either fertility rises or mortality declines (or the net balance of immigration turns up, which we can exclude as a possibility for the Europe of the 10th to 14th century). As historians tell us, it was above all a sustained change in mortality that mattered: while fertility as measured by crude birth rates was and remained high all throughout, mortality went down significantly – thus turning a society that was young, but stagnating into one that was still young, but growing.

Why did mortality decline? In pre-modern times, a high mortality was mostly the consequence of plagues and diseases. These, in turn, had the most devastating effects whenever the people hit by epidemics were in a poor physical condition, and that was mostly the case when hunger, famine and malnutrition prevailed. Therefore, given the essentially rural character of early feudal society, the main cause of a decline in mortality was almost certainly an improvement of agricultural productivity that led to a better food supply and less hunger and famine. Thus, in the end, the demographic take-off in the centuries under consideration is not really a separate phenomenon to be explained. Rather it is a consequence of the productivity growth in agriculture that set in motion a long chain of effects, among them an improved nutrition and health of the population.

So much for the demographics. As to economic performance in general, no straight and simple comparative numbers can be given to illustrate the presumed growth. A qualitative account must suffice, and that has to start – once again – with the productivity increase of manorial agriculture as described in section 2.2.1 of this manuscript. To be sure, a part of the improved food supply was consumed, and thus improved people's health and reduced mortality. However, a significant surplus remained. And this provided the necessary leeway for a process of trade integration and specialisation that set in at that time. Drawing a courageous parallel to the industrial revolution of the 18th and 19th century, the economic historian Robert Lopez has called this process “the commercial revolution of the Middle Ages”. At the heart of

this process, there are two forces that gained a new dynamics at the time: urbanisation and technical progress. To these we now turn.

2.2.2.1 Urbanisation

Nowadays, historians regard the rise of towns and cities in the Middle Ages as one of the big turning points in world economic development. And they have a good reason to do so: the cities laid the ground for a thorough commercialisation of life, which has not disappeared ever since in the western world, despite occasional setbacks and crises. And today, it is the urbanised “West”, notably Europe and North America, that sets the paradigmatic model for economic development and prosperity in newly emerging market economies everywhere on the globe. Remarkably enough, many developing countries have degrees of urbanisation that surpass anything the western world has ever reached. This is why the study of the economics of medieval cities and towns is not only important in its own right, but also a significant step towards understanding the forces that underlie modern Globalization trends.

What made towns and cities grow? The answer is uncontroversial: growth was due to immigration. People left the countryside to search for a better life in the city. In a very stylized form, the economics of the process may be described as follows:

- At the beginning, there was an increase of labour and land productivity in manorial agriculture. As a consequence, food supplies and health standards improved, mortality decreased, the rural population grew, and so did the surplus that was generated by rural economic activity.
- Given the fact that both productivity and population grew, it is not clear whether and, if so, to what extent the expansion actually improved the rural living conditions on a per capita basis. It did, however, broaden the leeway for activities outside agriculture, notably commerce, crafts, manufacturing and services.
- Economically, these activities concentrated in locations (“towns”) where “markets” for whatever products and services could be most efficiently organised. Physical proximity to agriculture was not required anymore as long as enough short-distance trade could develop to carry the produce to the city markets.
- Labour followed this urbanisation trend. Or it triggered it to the extent that it became itself entrepreneurial in the new urban activities. In any case, peasants left the land in

large numbers and turned into better-paid town-dwellers. In doing so, they could also leave behind their serfdom and become freemen, an additional incentive to migrate.

- Once towns and cities expanded, the growth dynamics gained momentum and fed on itself allowing further gains to be reaped in specialisation all over the economy, from commerce and crafts down to agriculture. Note also that agriculture itself began to diversify, with the distance to urban markets determining the crop to be grown.

So much for a stylized picture of urban growth. Naturally, the process also led to the foundation of many new small towns that had particularly favourable conditions for serving as central market places of a region. Thus an ever more intricate network of trade-linked economic activity developed in space. All this looked much different from the patchwork of manorial villages that had existed before the commercialisation had set in.

It is important to realize that this dynamics of urbanisation had a bright and a dark side. The bright side was the economic productivity of specialisation, which allowed a certain prosperity to develop – the first major silver line at the economic horizon after the Dark Ages. And with it came a creativity and originality in civic organisation, urban culture, public buildings and architecture, which are still today much admired by tourists, who walk through the medieval centres that are to be found in many otherwise modernised European cities.

The dark side were the living conditions in the towns, which must have been repulsively unhygienic by any modern standard, but maybe also by the standard of classical Rome with its remarkable achievements in sanitation and water supply. The typical European town of the Middle Ages had nothing like that. And it is therefore not surprising that the scourge of epidemics with a high death toll was a frequent visitor in medieval urban environments. This is the main reason why, typically, the towns had a negative demographic balance in terms of births and deaths. However, this did practically never lead to a sustained shrinkage of population size because, whenever plague or disease claimed many lives in town, the numbers quickly recovered through massive immigration from the countryside, where the demographic balance was typically positive. In this sense, as Carlo Cipolla has put it, the medieval cities were also large graveyards.

It is important, however, to realise that this dark side of medieval city life did never really destroy the attractiveness of the towns as a destination for potential migrants. Whenever labour scarcities emerged in the cities, people took the opportunity to fill the slots. In modern economic terms, we would speak of an apparently high “option value”: people migrated because

they expected an income stream high enough to more than compensate for the probability of professional failure, job loss or even death in a filthy urban environment with a high risk of falling victim to fatal plagues and diseases. That this option value was high enough to make people migrate at obviously high risks is an economic indication that, after all, the expected income stream must have been relatively high - compared to the available alternative of remaining a poor peasant-serf on a typical low-risk manorial estate.

Note that there are many parallels here to the situation in 19th century industrialising Europe, when cities like Birmingham, London and Manchester in England and Berlin, Paris or Vienna on the continent expanded rapidly and, at least for a while, were unable to provide a decent municipal infrastructure to improve living conditions. Similarly, the present situation of urban agglomerations in developing countries has an analytical similarity to medieval European times: e. g., Indian cities like Calcutta, Madras or Mumbai continue to grow through migration from the countryside although the living conditions in large parts of these gigantic metropolitan areas involve much more risks for life and health than those in rural areas.

Besides these parallels between the Middle Ages and modern times, however, there are also major differences. The most important one is the fact that, from the 10th to 14th century, urban growth took place against the background of a rural world of feudalism. Thus, for most people, migrating from the countryside to the city meant moving from serf- to freedom. It also meant leaving a hierarchical order behind and replacing it by a world, in which voluntary horizontal arrangements dominated. In fact, it is characteristic that, within the growing cities, free associations of people with common interests began their long and proud history that reaches up to the present. The gild, the confraternity, the university and the commune itself were all co-operations among equals, far away from the rural world of lords, vassals and serfs.

In a sense, the urban “res publica” of Roman times was back on stage. And it is typical that the politics and administration of urban municipal life in the Middle Ages has much more in common with classical Roman practice than with the contemporaneous feudal structures in the countryside. E. g. in Venice, which had no Roman past, municipal decision-making came about through assemblies, councils and even a senate, where patrician families made politics – very comparable to Republican Rome. Admittedly, Venice may be an unusual example, but even more humble cities like many towns of the Hanseatic League did have bodies of collective decision-making that allowed for a strong influence of leading citizens, and not just the will and whim of a singular person.

To be sure, the mere emergence of “horizontal” urban institutions did in no way remove the economic rationale for the implicit contracts that underlay the hierarchies of feudalism. In rural areas, they worked and remained intact for a long time. After all, as we have seen in section 2.2.1, they were appropriate answers to an extremely high degree of isolation and insecurity in a thinly populated rural world. In a densely populated urban world, however, that isolation vanished, and given the high degree of communication and control that could be organised at short distances within a city, the provision of security as a major public good could again be organised on a municipal level.

Hence the Middle Ages became increasingly characterized by a remarkable dichotomy, which had both an economic and a political dimension:

- Economically, it was the urban world of commerce, trade, crafts and manufacturing versus the rural world of manorial agriculture.
- Politically, it was the urban world of freedom and cooperation among equals versus the rural world of hierarchy, subordination and serfdom.

Note that the boundaries between these two worlds were sharply visible in the form of the city walls that separated the inside world of commerce and freedom from the outside world of agriculture and serfdom. Note also that the importance of these boundaries were clearly recognized by contemporaneous observers; and they were often commented on with value-laden judgements, depending on where the sympathy of the observer lay. E. g., moralists like Dante liked to complain about city life and decadence whereas urban citizens loved to boast about the economic and cultural achievements of their home towns.

To be sure, there was a lively interaction between towns and countryside, and it grew as time went by and urbanisation proceeded. Depending on the power balance between feudal lords and the cities, there were very different types of interactions in different regions. E. g., the powerful Italian cities of the late Middle Ages and Renaissance times like Florence and Venice exerted a strong political influence in their hinterland whereas, in medieval Germany, it was the other way round, with feudal lords meddling in city politics. By and large, however, the dichotomy did survive until modern times, though clearly in a much more attenuated form.

It is important to realize that this dichotomy marks a major difference to the Roman empire – and, for that matter, to other ancient empires. Towns in all these empires were not autonomous worlds of freedom and cooperation. Rather they were political and administrative centres, which governed all the rest of the empires. This is a big difference, and it points to the

core of the role that the medieval city was bound to play in the further history of the western world of commercialism or, for that matter, capitalism. In medieval towns and cities, merchants and craftsmen began to exert a major influence on politics, be it individually or through guilds or other associations. Thus, maybe for the first time of history, genuine business interests entered politics and thus laid the ground for policies that were conducive to commerce and trade as well as crafts and manufacturing.

Note the subtle, but important differences to the Roman empire, which we may summarized in a provocatively pointed form:

- There, in Rome, a commerce-friendly legal system and a remarkable infrastructure of roads were not more than incidental by-products of a policy that, at heart, aimed at Roman glory, grandeur and military might. Here, in the medieval city, policies are systematically lobbied to the direction of commercial interests.
- There, in Rome, commerce was considered as an unavoidable evil to fill the pockets of any gentleman to make him able to pursue the nobler goals of public life and discourse. Here, in the medieval city, the patricians are often themselves highly respected merchants and to a lesser extent craftsmen and manufacturers.
- There, in Rome, the city heart was the forum, i. e. the location for political discourse and public presentation, with markets being adjoined, but only in second place. Here, in the medieval town, the city heart is the marketplace, with the town hall being closely attached – like an authority that regulates commerce rather than a centralized administrative powerhouse.

From its beginnings in the 10th until the late 13th and early 14th century, the urban expansion of the Middle Ages could go largely undisturbed. In economic terms, it did not meet any major bottlenecks or run into diminishing returns, which could have set serious limits to growth. Of course, the expansion of population and the towns required more virgin land to be put into use, but for a long time there was plenty of it. Forests were cleared to make room for settlements and cultivation, and more often than not, the additional land was of very good quality and could be most efficiently put to use in the more specialised pattern of agriculture that began to emerge.

All this began to change when in the late 13th and definitely in the early 14th century, the expansion did eventually arrive at a variety of bottlenecks. Historians tell us that two of their kind stood out in importance:

- Agricultural expansion ran into diminishing returns. The remaining lands to be cleared and cultivated were not anymore of a good quality. The frontier of agriculture had been pushed further and further into marginal territories with poor soils and climatic conditions. As a consequence, it became ever more difficult to provide sufficient food for a growing population. As far as we can know from poor historical records, food prices rose and real wages fell, thus wiping out a part of the progress in living standards that had been achieved during three centuries of expansion. Note that, even in the expansionary period, the fast population growth put a break on the increase of real wages so that certainly not the full-scale productivity gains could be reaped by labour. To be sure, the many impressive, highly ambitious building projects of the Church and the richer cities do point to a very unequal distribution of income all throughout the period, with large funds at the disposal of a few privileged decision makers (as, e. g., the church) and labour being in relatively cheap supply to work on the construction sites of cathedrals, town halls and the like.
- Population density in the cities reached a level that was not sustainable without a fundamental improvement in sanitary and hygienic conditions. Apparently, a dangerous situation of overcrowding gradually built up. It finally erupted with a series of diseases and plagues, which culminated in the famous pandemic of 1348-51 that is known under the horrifying name “Black Death”, which killed about 25 million people. To be sure, minor epidemics had reached different parts of Europe before and others were to arrive in the decades thereafter. Nevertheless, the Black Death figures as a watershed that signals the end of a long all-round expansionary period of history and ushers in a new phase with different characteristics. Most notably, the death of about one third of Europe’s population – in some areas even much more – led into a renewed phase of labour scarcity, which turned out to be quite different in economic characteristics from the early feudal times. We will come to it in section 2.2.3 of this manuscript.

Whatever the reasons in detail, the great expansion had come to an end in the mid 14th century. As we shall see later, urbanisation proceeded nevertheless, but at a considerably slower speed. Therefore it makes sense to pause for a few numbers. How far had urbanisation gone in the decades before the Black Death?

As usual for medieval times, the numbers are estimates, but for that matter, they are much better estimates than those for large medieval regions or countries because cities and towns have left more paper tracks and records to be analyzed by astute historians. In the late Middle Ages and early modern times, there were even regular censuses held in larger cities, and – apart from delivering relatively hard facts and data – these censuses are themselves an indicator of a growing rational awareness of how important the knowledge of the population is, if only to assess the city's economic power and fiscal capacity for taxation purposes.

All in all, the numbers tell a story that is simple and clear: there was a first significant urbanisation, at least in the most advanced region of Europe, which was northern Italy, but it did not nearly reach the level of Roman times. The four by far largest European cities were Florence, Milan, Venice and Paris, the only non-Italian one, all with populations around 100.000 people. There is no indicator that could show that any other city came close to this. The next in the row were Bruges in Flanders with 35.000 and Padua in Italy with 30.000 people. The only German city that might have reached a population of 30.000 is Cologne, but we cannot really know because reliable numbers for this early period are not available.

To sum up, there is not really any region in the period 1000-1350 that was “urbanized” in the sense that it had enough big cities, which could have been the basis for a thoroughly mercantile society with a very strong emphasis on commerce, trade and specialisation. There is only one debatable exception, and that is northern Italy. It should therefore not come as a surprise that our examples for long-distance trade networks that will be discussed in sections 2.2.4 and 2.2.5 of this manuscript will focus on a set of special cases, part of them Italian ones, which are not representative, but of great exemplary importance.

2.2.2.2 Technical Progress

As we have seen in section 2.2.1, major innovations in agriculture during early feudal times laid the ground for slow, but sustained productivity growth, which in the end paved the way for new market-oriented economic activities in an urban environment. Naturally, technical progress was not confined to agriculture. To the contrary, the expansion of commerce, crafts and manufacturing led to a new and broader wave of innovation that allowed to reap gains in productivity in various branches of an increasingly differentiated division of labour.

A major breakthrough was the widespread use of water and wind mills for generating energy. In a sense, this was an altogether revolutionary step forward because it meant that, for the

first time in history, mankind could extensively rely on an inanimate power source. With the benefit of hindsight, it appears to be not so surprising that, once inanimate power was systematically tapped, it also became part of human imagination about further improvements in its efficiency, thus opening all the way up to the steam power that ushered in the industrial revolution in the 18th and 19th centuries.

Remarkably, water and wind mills had been invented outside Europe. The water mill was a traditional Chinese technology, which is documented in Europe not before the 9th century, but then in a wide range of industrial activities: firstly in Picardy for grinding malt in beer production, then from the 11th century in today's England, France and Germany for fulling cloth in the nascent textile industry, from the 12th century in Styria, Normandy, southern Sweden and in Moravia for iron production, and from the 13th century for the operation of saws and the manufacturing of paper in many different towns and villages. In turn, the wind mill had been invented in Persia in the 7th century AD, but it emerged on the European stage not before the 12th century, possibly as a technology import from the Crusades. However, due to some ingenious technical adjustments, the European type of mill was from the beginning more powerful than the Persian original (and, of course, than water mills). In fact, it could generate as much as the power of 20 to 30 horses, but of course remained confined to those regions that had appropriate climatic and geographic conditions.

Beside the mills, there were other major technical innovations in crafts, manufacturing and transport services. In textile manufacturing, the vertical loom was introduced in Champagne and Flanders in the 11th century. As a replacement for the traditional horizontal loom, it is estimated to have pushed up labour productivity by a factor of five (and besides improved the quality of the woven fabric). In the 12th century came the inventions of the compass, the water-clock (for measuring movements of the ship), naval charts and trigonometric tables for mathematical navigation, which all improved the rate of capacity utilisation of ships as a major part of the capital stock in maritime transport. E. g., in the Mediterranean, a ship was then able to complete two round-trips a year instead of just one because navigation could be continued in winter, despite cloudy skies and bad weather. The 14th century saw mechanical inventions like, e. g. firearms, locks for canals and the clock, which allowed for the first time to measure precisely the exact length of labour units, a matter that was to have paramount importance for the future organisation of economic life.

Technical progress is one thing, the diffusion of new technical knowledge another. Naturally, there is very little hard evidence on the interregional flow of ideas for a time when little was documented anyway. Historians agree that technology spread mostly in an embodied form,

i. e. through the migration of people. In this respect, examples abound and deliver a clear-cut pattern: skilled workers were attracted to those regions where they could find plenty of the complementary factors of production that they needed. A major case in point was the colonisation of the wide plains of eastern Europe by peasants from the West. In the 12th and 13th century, the eastward movement, the so-called “Drang nach Osten” (“push to the East”) by the German knights can be considered as an early example of a massive development policy, which used generous grants of virgin land as a boon to pull skilled labour from their western homelands into the thinly populated north-eastern regions of Europe.

Remarkably, this type of active policy of immigration and development was not confined to those authorities who administered wide empty spaces. To be sure, towns and cities did essentially the same thing, even when they had already reached a relatively high standard of living and a well-established crafts-tradition. In fact, the first and most active in the competitive game of attracting skilled labour were exactly those northern Italian communes that had already established a quite high degree of specialisation in manufacturing. E. g., in a well-documented case, the city of Bologna granted massive tax privileges and subsidies to artisans who were ready to move to the city and to set up business in the nascent wool and silk industry. This happened as early as 1230-1. As a matter of fact, the incentives were highly effective: they attracted 150 artisans to the city, which at the time had no more than 25.000 inhabitants, which amounts to a population increase of 0.6 %, all in terms of highly skilled and entrepreneurial people. And, economically, the whole policy paid off handsomely: Bologna’s textile industry boomed thereafter and remained a major pillar of the city’s economy for centuries.

To sum up, it is fair to say that there were many technical innovations in the later Middle Ages. And that public authorities knew exactly how critical the import of technical knowledge and skills was for economic development. And that quite a few of them acted accordingly. Hence nothing could be more wrong than the traditional, but long since refuted view that the Middle Ages lacked any dynamics of human ingenuity. May be that this odd view is nothing but a long-standing prejudice that was born and cultivated by Renaissance intellectuals, who had a tendency to denigrate their ancestors of previous centuries as indolent ignorant people who had not yet been enlightened by the rediscovery of the classics and their ideas. Today we see much clearer that, at least in terms of economics and technology, there is much continuity in progress between the “humble” Middle Ages and the “proud” Renaissance times.

In one respect, however, the Renaissance critics may not have been all that far off the mark: most medieval innovations were not really big inventions, but rather modest adjustments and

improvements to the specific needs of the time. In fact, it is not at all an outstanding originality of human thinking that is characteristic for Middle Ages. Rather it is a widespread curiosity and open-mindedness, and thus an extraordinary readiness to absorb ideas that had already existed for long, and to work hard to apply them productively.

On top of this absorption capacity, there is something else in the later Middle Ages that turned out to be of great importance for the future: a real fascination with mechanics, which then reached a first peak in Renaissance times with Leonardo da Vinci's numerous visions of fantastic machinery. Other than in Ancient Greece or Rome, it was not anymore philosophy or artistic harmony or law and justice that inspired the imagination of mankind, but machines and mechanics. The economic historian Carlo Cipolla regards this fascination as a full-scale "mechanical outlook". He holds that it was this outlook that made medieval man search for mechanical solutions wherever possible and feasible, be it to generate power in mills or be it to measure time by clocks. The "machine" as a thoroughly mechanical device thus became an aspired end in itself because it promised man to dominate nature – and to leave the state of harmony with nature that had been the ideal in ancient world of Greece and Rome.

Remarkably, Carlo Cipolla goes even a step further. He recognises the mechanical outlook in the medieval cult of the Saints, which contrasts sharply with the values of the Homeric world. Whereas, in Homer's oeuvre, the hero who challenges the harmony of man and nature receives awful punishment and ends miserably (think, e. g., of Daedalus or Prometheus), the moral message of the Saints cult points exactly in the opposite direction: Saints are human beings who overcome the forces of the world, they harness nature, they perform miracles, i. e. they do things that appear to be mechanically impossible, things that humbler humans only dream of. And they receive admiration and adoration for it – and a place in paradise. Could there be a greater incentive to push forward the limits of the realm of mechanics in this world?

2.2.3 The Time of Scarce Labour

The famines and plagues that culminated in the Black Death of 1348-51 led to a drastic shrinkage of the population in Europe. About 25 million people died. With the benefit of hindsight, we know that for the rest of the Middle Ages and Renaissance times, Europe never again reached its pre-plague population peak of 80 million people. Gradually, the population grew again, but it is estimated at no more than 60-70 million at the end of the 14th century. This is why, for a long time, historians regarded the period 1350-1500 as a period of economic depression. Add to this the fact that it is also a period of long-drawn out military con-

flicts in major regions of western Europe, notably the Hundred Years' War (1337-1453) that devastated parts of France, and the War of the Roses (1455-85) that followed in England. It is then in fact tempting to conclude that it all amounted to a serious and dramatic backlash against the background of the "Great Expansion" of 1000-1350.

As a matter of fact, this sombre view of the period is no longer accepted by most historians. The modern view is that the expansion and the contraction should be viewed not as a cycle of rise and decline, but rather as a pattern of growth and structural adjustment to temporary growth limits and bottlenecks. This modern interpretation has paradigmatic importance as we shall see in the further proceedings of the lecture. Until in our days, the world economy has been repeatedly subjected to bottlenecks and limits to growth. As examples, we may think of the so-called oil crises in the 1970s and 1980s or, two centuries earlier, the acute scarcity of energy due to the destruction of forests and the concomitant soaring prices of charcoal. In these cases – as in others – the world economy had to adjust, and where that happened successfully, there tended to emerge a new growth dynamics in due course. It is therefore important to have a somewhat closer look at the basic economic logic of what happened in the period 1350-1500. After all, it is the first major period in history when a sustained expansion apparently ends in some sort of disaster, which – in the very long run – proved to be the basis for a resumption of growth.

We first look at the most important stylized fact of the time: the scarcity of labour. Anecdotic evidence abounds that the famines and plagues of the 14th century did not only drastically reduce the size of the population, but ushered in a period of acute labour shortage. Apparently, economic pressure followed demographic trends. Note that this is not self-evident because a shrinking population means not only a decline of factor supply, but also a decline of demand for goods and services to be produced. It is a first – though rather off-hand – indication of the vitality left in the economy after the Black Death that labour turned out scarcer than before.

How pressing was the labour shortage? Fortunately, two British economic historians – E. H. Phelps Brown and Sheila Hopkins, have been able to deliver some solid quantitative evidence, at least for England. They calculated the development of monetary wages of a building craftsman and a labourer in southern England from 1264 to 1700. FIGURES 3 and 4 present their results graphically, FIGURE 3 in nominal and FIGURE 4 in real terms. The message is unambiguous. FIGURE 3 shows that, from about the mid 14th century onwards, the nominal wage rose sharply. In fact it roughly doubled for the craftsman and almost tripled for the labourer – thus incidentally indicating a narrowing of the skill differential, although one

should probably not read too much into these numbers in detail. FIGURE 4 shows that the nominal rise did more or less fully translate into a real one. This indicates that the standard of living of workers indeed rose, and quite substantially so. Note that this real increase did not continue beyond the end of the 15th century: from that time until about 1600, there was still a stepwise nominal increase (FIGURE 3), but that was more than fully neutralized in real terms by the sharp price increases that occurred in the inflationary 16th century. Not before the 17th century does the rise of real wages resume.

True enough, this evidence covers only a very small part of Europe, but historians agree that its major message can be generalised. After all, there are many supplementary pieces of evidence for other parts of Europe that point in the same direction. Remarkably, the indicators also suggest a quite dramatic change of income distribution at the time: while wages of craftsmen and labourers were rising, rents and interest income declined. And as far as this evidence goes, the decline must have been quite drastic. Note also that, within the cities and towns, evidence on consumption data suggests that the income gap between the rich merchants on one side and craftsmen and workers on the other narrowed: gradually, luxury goods were not anymore exclusively confined to the most wealthy classes, but became available on a somewhat broader basis.

From all this it becomes evident that the period 1350-1500 was certainly not a time of massive all-round decline. The picture is much more differentiated. Apparently, market forces led to a substantial rise in real wages, and that is likely to have led to structural adjustments that may have been advantageous in the long run. After all, a state of labour scarcity is conducive to the introduction of labour-saving techniques, which in the end may raise productivity and living standards everywhere. In fact, there is no evidence that the speed of technical progress may have slowed down after the Black Death. To the contrary, the impression is one of essential continuity in progress.

We have arrived at a point of the lecture where it is most useful to briefly set up a highly stylized aggregate production model of the medieval economy (or, for that matter, for any economy that we are dealing with in this lecture). Let us postulate with bold simplification that the European economy in the period 1000-1500 can be described by a standard neoclassical production function $Y = Y(L, T, K, X)$, with Y defined as output or income (in real terms), L as labour input, T as land input (T for “territory”), K as the capital stock in use and X as all the rest that affects aggregate productivity in an economy, from the degree of specialisation to the level of technology. Note that changes in X reflect all those manifold developments and trends over time that cannot be subsumed under a change of one of the other factor inputs.

We assume, that all inputs have a positive marginal productivity, i. e. formally $\partial Y / \partial L, \partial Y / \partial T, \partial Y / \partial K, \partial Y / \partial X > 0$. We further assume that, for L, T and K, the (positive) marginal productivity declines with the extent of the respective factor input, i. e. formally $\partial^2 Y / \partial L^2, \partial^2 Y / \partial T^2, \partial^2 Y / \partial K^2 < 0$. Economically, this means that the productivity of the inputs L, T and K declines with the extent of factor input if the input of all other factors is held constant. This is the “law of diminishing returns”, which has a fundamental intuition on its side, especially for the purpose of analyzing economic history in the long run. After all, it is hardly conceivable to imagine a world in which the input of ever more labour, land or capital would generate an ever constant or even increasing additional output if the input of all complementary factors remains unchanged. In fact, if something like this were observable, we would have to search for a “hidden factor” that might solve the puzzle. In fact, X may be considered as some such hidden factor.

Note that we make no assumption on diminishing returns of X. We have good reasons for this: other than L, T and K, the variable X is not really a physical input, but rather a “catch-all”-variable that may affect the productivity in the economy. A priori, we have no idea of whether, say, the level of technology or the gains from specialisation underlie anything like a tendency to increase or decrease with the level already reached and at given inputs of the other factors L, T and K.

Note also that we assume that all factors are complements – in the sense that an increase in the supply of one factor increases the marginal productivity of all others. Formally this means, for the exemplary case of an increase in K: $\partial^2 Y / (\partial L \partial K), \partial^2 Y / (\partial T \partial K), \partial^2 Y / (\partial X \partial K) > 0$. Note finally that we assume that factors are paid their physical marginal product (in real terms) or their value marginal product (in nominal terms).

So much for the basic assumptions of our model. We will use the model (or variants of it) at different points of the lecture. We do so for purely illustrative purposes. This is why we do not bother to make any further more precise assumptions on the model that may be necessary to keep its properties under control. For our purposes, we just need the basic logic that comes out of it – and not any more exact specification.

In terms of this simple model, we can now describe the whole period 1000-1500 in a simple stylized way. In doing so, we divide the period into three crudely defined sub-periods, the first one running from 1000 to 1300, the second one from 1300 to 1350, and the third one from 1350-1500. The first one may be called the time of “all-round expansion”, the second one the

time of “limits to growth”, and the third one as the time of “structural adjustment”. We describe these sub-periods very briefly in the following paragraphs.

In the first period, all factor inputs go up: demography makes L rise, the clearing of forests for agriculture and settlements increases T , profitable investment in machinery for agriculture as well as commerce and manufacturing pushes up K , and the rising level of technology and the increasing intensity of trade make X go up. It is a situation of “all-round” growth: L , T , K and X all grow, and so does, as a consequence, the economic performance of the economy as a whole. However, it is not clear who or which of the factors of production gains most or least or even loses in this process. That depends on a possibly subtle balance between quantities and prices of factors. In particular, the balance between labour on one side and the owners of land and capital on the other is not clear. Is the growth of the labour supply so fast that all gains in marginal labour productivity realized at a given level of employment are “eaten up” by the expansion of employment that leads to a declining marginal productivity and real wage? Is the expansion of the supply of land and capital so large that all gains in the rate of return on a given supply are “eaten up” by the additional supply on the market?

FIGURE 5 illustrates the main forces at work for the case of labour in a standard graph of a production function in labour/output-space. At a given level of employment L^* , the marginal productivity of labour is pushed up over time so that there is a “potential” for a real wage increase. Whether this potential can be realised depends, however, on how much the labour force increases – and thus pulls down the marginal productivity and the real wage along the curve that represents the production function.

Do we have clear evidence on the balance of forces for the period 1000-1300? The answer is no. We know that the economy and all factor supplies grew and that specialisation increased and technology progressed, but we do not know much more. If anything, from the juxtaposition to the later period, we just guess that the balance is not unlikely to have been tilted against wage-earners and in favour of land and capital owners.

The second period is the one when, gradually, the economy begins to touch the limits of growth. Formally, this may be viewed as a further growth of the labour supply, but at the same time some inherent brake on the supply of (good-quality) land and on the options of further extending the capital equipment and intensifying specialisation in overcrowded cities. A kind of “Malthusian trap” emerges – named after Thomas Malthus, a British demographer and political economist of the late 18th and early 19th century, who claimed that there is a definite limit to population growth in an agrarian society (as the one of his day) because,

once poor-quality marginal lands are put to use to feed the population, wide-spread starvation follows and brings back the size of the population to the lower sustainable level. In fact, in a somewhat broader sense, this is exactly what happens at the time of the Black Death. Famine and plague take their toll and drastically reduce the labour force to a level that is again sustainable.

The third period is the one with the new equilibrium. The population has shrunk to a level that is sustainable. At the same time, it is now labour that is scarce and expensive. FIGURE 6 describes the situation graphically. The Black Death has pushed the economy back to a level with a higher marginal productivity of labour – and, as a consequence, a higher wage rate. The owners of land and capital, however, are the losers. They find themselves deprived of the complementary factor labour that used to raise their own productivity and remuneration. In a dynamic context, which goes well beyond the graph, it is the (entrepreneurial) owners of land and capital who have now a strong incentive to introduce labour-saving techniques and to fuel productivity growth anew.

2.2.4 Maritime Trade

Urbanisation opened the door to a new wave of integration. In fact, the period 1000-1500 was essentially a time of uninterrupted growth of trade and commerce, with the Black Death not really leading to a trend change. After all, virtually all large cities were centres of crafts and manufacturing, and apart from the short-distance trade that developed around the towns with mostly imports of food and agricultural produce in return for exports of manufactures and services, a long-distance network of trade routes began to emerge, stabilize and expand. Of course, transport costs were high all-round, but they were clearly much higher on land than on water. This is why, for a long time, it was to be the dynamics of maritime trade that dominated the picture of what may be called long-distance inter-city commerce.

How did this trade look like? Where took it place? What was its economic significance? To answer these questions, it makes sense to start from the distribution of large cities in late medieval Europe (FIGURE 7). Note the categories used: a “very large” town is one with at least 25,000, a “large” one with at least 10,000 and an intermediate one with at least 2,000 inhabitants. Three points are noteworthy at the outset:

- There were no more than 20 “very large” cities in Europe at that time. These “very large” cities were distributed very unevenly over the continent, with major agglomerations being in northern Italy (five: Bologna, Florence, Genoa, Milan and Venice) and

in a quite narrow triangle of north-western Europe, i. e. in Flanders (three: Bruges, Ghent and Ypres) and “around” Flanders (another three: Cologne, London and Paris).

- “Large” towns – the category in the FIGURE 7 below “very large” ones – tended to be clustered in and around the two core regions of urbanisation. “Intermediate” towns – the category further below – were much more widely scattered, with most of them to be found in western and central Europe including the whole of Italy, but excluding the Iberian peninsula and the British Isles.
- In general, towns and cities were most likely to be found along rivers or in the vicinity of them. Clearly, the big river systems of continental Europe served as the cradle for medieval urban growth: Po, Rhone, Garonne, Loire, Seine, Thames, Rhine, Elbe and upper Danube, and – still on a very modest scale – Odra and Vistula, they all and their tributaries played a part in providing not only rich soils in alluvial plains, but also navigable water arteries that lowered transport costs for commerce and trade.

Remarkably, there was not really a very powerful concentration of cities along the shores of the ocean, i. e. the Mediterranean, the Atlantic, the North and the Baltic Sea. True enough, the Mediterranean could boast with six “very large” coastal cities (Barcelona, Genoa, Naples, Palermo, Venice and Constantinople), but that was not comparable anymore to the urbanisation of ancient times, and it was – more than ever before – concentrated on Italy. In turn, the growth poles in north-western Europe relied more heavily on a combination of river and sea transport so that big cities right on the coastline were anyway the exception rather than the rule.

It is this simple observation that leads us straight to a very important characteristic of the link between urbanisation and the expansion of long-distance trade in the Middle Ages and in later periods. As towns grew and developed highly specialised crafts and manufacturing industries, there was also a growing need to export their wares in return for imports of agricultural produce, raw materials and resource-based products. True enough, this trade structure was already known in ancient Greece and Rome, but it gradually reached an altogether new quantitative dimension as well as breadth and depth in the course of the medieval “commercial revolution” and in subsequent Renaissance times.

FIGURE 8 gives a first clue why this may be so. It shows the European centres of textile production in the 13th century. On first glance, it looks almost like a replica of FIGURE 7, and in a sense it is. Sure enough, the big urban agglomerations of medieval times in northern Italy

and in and around Flanders were also the centres of the textile industry, which was at the time by far the most important branch of manufacturing. After all, the most basic needs of mankind are food, shelter and clothing, and once technical innovations like, e. g., the vertical loom had sharply raised labour productivity, there was ample scope for growth in textile output. As the history of clothing habits tells us, the new supply met a powerful wave of differentiated demand. E. g., the custom of wearing shirts and underwear began to spread in the Middle Ages and created a new demand for woven fabric that had not existed before. In fact, the very word “shirt” (in Latin : camixia) is of medieval origin; it did not exist in ancient times, which shows that an altogether new dimension of needs and wants was opened by the new technologies of textile production.

Clearly, there was ample scope not only for more textile production, but also for more trade in textiles. To be sure, textiles share three characteristics that make them particularly suitable for long-distance commerce.

- Textiles are light, i. e. in more precise economic terms, they typically have a high value added - and price - per unit of weight. Among all traditional types of manufactures (others are, e.g, glass and metal wares), textiles are probably the lightest. As a consequence, transport costs are low, which facilitates trade.
- Textiles are produced highly labour- and not land-intensively so that they typically qualify for the type of trade that is described in standard textbooks as following the so-called Heckscher-Ohlin-model, i .e. trade according to factor proportions (e.g., exports of labour-intensive textiles in return for imports of land-intensive grain). As we shall see below, this was in fact a major pattern of trade that could be observed.
- Textiles are enormously rich in variations of type, quality and style – ranging from coarse knitwear and woven fabric up to delicate silk garments. Thus, in principle, many specialised producers can profitably engage in trade if they develop their own strong “commercial niche” – a classical case for monopolistic competition in so-called intra-industry trade that is covered by any modern textbook on trade.

Given these conditions, it is not surprising that the textile industry played a kind of strategic role in the up-coming European long-distance trade networks, both on water and overland: textile manufacturing did not only make the many urban centres realise substantial productivity and income gains, but it also allowed major seaports (and land trade fairs) to take over a

most profitable role in the commercial chain that developed. To these we now turn our attention.

In maritime trade, the economic rise of some seaports was spectacular indeed. At no time in history before or after did single commercial cities take such a prominent position as in the late Middle Ages. This is exactly what makes them so interesting as early examples of what may be called global capitalism: in a still feudal rural environment, they ventured into a new world of commercial sophistication and economic prosperity.

To be sure, this success was partly a reflection of the weakness of the central political power, which was nowhere strong, but especially weak in the Holy Roman Empire of German nations. There, an emperor formally ruled over most of central Europe - from the German shores of the North Sea and Baltic Sea to the northern Italian coastlines of the Mediterranean (see FIGURE 9). However, his actual political power was quite limited at any time and virtually nil in some periods and some territories. It is therefore no coincidence that we observe the most remarkable careers of seaports in or close to the outskirts of this empire: the rise of Venice and Genoa in the South and the Hanseatic League in the North. Still then, the driving forces that fuelled this rise were not political, but economic ones. To these we now turn.

2.2.4.1 Southern Europe: Venice and Genoa

Right from the beginning of the commercial take-off in the 10th century, Italian seaports were particularly well-placed to take advantage of the new economic opportunities. Four of them – Amalfi, Genoa, Pisa and Venice – had a most promising early start. But, for a variety of economic, political and military reasons, two of them – Amalfi and Pisa – ended in less prominent positions. The two others, however, rose up to the European economic top and left deep traces in history: Venice and Genoa. Both cities had two distinctive geographical advantages in common: (i) they were part of or not far from the big centre of the urban rise of manufacturing in northern Italy; (ii) they had close and easy access to major Alpine Passes that linked them to the main north-south routes of land trade in central and northern Europe. Both combined these advantages with their – typically Italian – integration into the more traditional trade network of the Mediterranean, which stretched from the Atlantic Ocean in the West to the Black Sea in the East.

Venice is the older and the more stable one of the two. The lagoon city in the Adriatic Sea had no Roman past, and its early development in the 6th and 7th centuries AD is shrouded in

legends. As a political entity, it was practically independent from the 9th century onwards. Geographically, it found itself ideally placed between the new growth poles of the northern Italian plains and the Byzantine empire, which at the time was still much more prosperous than Europe. Constitutionally, the city of Venice was a republic with a strongly aristocratic flavour: on top the Doge (in English: "Leader"), selected for life-time from one of the great merchant families who also figured prominently in the various councils that, in a kind of pyramid structure of power, had important advisory functions, from the high "Ducal Council" over the "Senate" and the "Great Council" down to the "General Assembly". Whatever the details of this power structure, it remained remarkably stable over time and persistently dominated by commercial interests – at first those of the powerful merchants and then also of (export-oriented) industrialists.

Politically, the Republic of Venice delivers an excellent example for a consistent mix of policies that subordinated practically everything to the commercial success of the city and the seaport. By extremely astute political and military manoeuvring, Venice obtained massive commercial privileges all over the Byzantine empire (including Constantinople) as early as 1082, in exchange for naval support against Norman aggression. In fact, the terms of the so-called "Golden Bull" that guaranteed these privileges were such that Venetian merchants were better treated than Byzantine merchants in their own homeland, thus constituting one of the rare cases in history of positive discrimination of foreigners. More than a century later, in 1204, the 4th crusade gave Venice a unique opportunity for a masterly use of political diplomacy, commercial clout and military pressure to enlarge its power. The outcome was a partition of the Byzantine empire: parts of it were incorporated in what from then on could be called a genuine Venetian maritime empire that consisted of many seaports and narrow, but long stretched-out coastal regions and islands (including Crete) in the eastern Mediterranean (See FIGURE 10).

Note that, in its organisation, the Venetian "empire" had more the character of a colonial system of trading bases than of a genuine empire as the Ancient Roman one with a highly centralized bureaucracy and army. Economically, there was anyway not much difference between the many Venetian trading bases inside and outside the empire, which together formed a closely-knit network of representations and entrepôts – all, in modern terminology, direct investments with the clear-cut strategic motivation of guaranteeing a smooth working of commercial activities at the location where the buying or selling actually took place. As a matter of fact, big seaports like Alexandria and Tyre, which were formally outside the empire, but served as Venetian important trading posts, had much greater economic importance for Venetian trade than many minor coastal cities within the empire itself.

Economically, Venice started its career almost exclusively as a trading city. For contemporaneous observers used to agriculture as the core of economic life, this appeared almost like a miracle. “These people do not plough, sow, or gather vintage … They come with their merchandise and buy grain and wine in every market place”. This is what a writer from Pavia in the Po valley said about the early Venetian way of life, apparently stunned by the fact that Venetians completely relied on the division of labour through trade, and quite successfully so because, visibly, it made them rich.

In fact, Venice was ideally located for trade, notably for transit trade, right at the centre of the commercial routes that connected the (still prosperous, but stagnating) Eastern Mediterranean with the (not yet prosperous, but growing) north-western European regions, via the (equally growing) northern Italian plains. In essence, the transit trade consisted of exchanging eastern wares like sugar, spices, porcelain and silken garments, which came from regions as far as China and India, for northern products like, e. g., woven textiles and various metals. The meeting points for this long-distance trade were at first the so-called Champagne fairs in northern France which were connected to Italy via the Great Saint Bernhard Pass (see FIGURE 11). Later-on, with the rise of the northern Hanseatic trade network, the trading shifted to the city of Bruges, a major Hanseatic commercial base where Venetian merchants were prominently present. In Renaissance times, the trading route moved eastward to the Brenner Pass, with Augsburg, Nuremberg and Leipzig becoming major inland trading centres that linked Venice with the Hanseatic seaports of Lübeck and Hamburg.

Note that, in political terms, the transit trade required a certain monopolistic dominance of Venice in the northern Italian plains so as to channel the commercial activity through the Venetian seaport itself (and not through other northern Italian cities like, e. g., Ferrara, which were also well placed in the lower Po valley). In fact, much policy effort was put into achieving this dominance and preserving it over time. To be sure, the policy was quite successful and, as a positive by-product, it laid the ground for a Venetian hinterland that, as a fertile and urbanised region, proved very viable in much later times when the stagnation of trade had set in and many rich Venetian citizens built mansions on the mainland and drew income from there. In a sense, the prosperous province of Venetia, which still exists today as a large north-eastern Italian “region” with cities like Padua, Verona and Vicenza, is the remaining long-term outcome of this policy.

Sure enough, trade in Venice was not only transit trade. It also consisted of imports that were destined for consumption and demand by the Venetians themselves, and of exports that

were produced right in the city. With the growth of the city's population and economy proceeding, this type of trade gained in relative importance:

- As to imports, it was above all grain that was needed in a big town without agriculture. In fact, once Venice grew into a city of about 100.000 people through massive immigration from its northern Italian hinterland, the security of grain supplies was a persistently public issue of major political and economic importance. This may be one of the reasons why Venice became the centre of the northern Italian market for wheat, where imports from Sicily, Egypt, Greece, the Balkans or even today's Ukraine were traded. Apparently, it was the city's own consumption demand, which was big enough to turn it, as a by-product, into a commercial centre for agricultural produce.
- As to exports, Venice was and remained a major producer of salt from the lagoons that was shipped in all directions. Over time, however, it was the growing manufacturing sector of the city that provided the biggest potential for high-value exports. Most prominent were glassware and other chemicals, textiles that were made from imported raw silk, as well as ships from its shipyards. Clearly, the growth of manufacturing and its specific pattern was closely linked to trade, but it is important to realise that, once trade ceased to grow in later centuries, Venice remained for a long time a prosperous manufacturing city.

At the peak of its economic development, say, from 13th to the early 15th century, Venice reached a mature stage of commercialisation which, in a sense, may be regarded as the economic maximum that could be achieved through trade integration and trade policy in an essentially pre-industrial rural environment, but in otherwise most favourable circumstances. Looking over the Venetian economy at that time, five major characteristics stand out in importance:

1. As a hub a long-distance trade, Venice grew into an "**international" commercial centre**". Foreign traders were actively invited to have trade representations in the city, the most famous one being, from 1228 onwards, the "Fondaco dei Tedeschi" next door to the market at the Rialto bridge, i. e. the "German trading house" where merchants from North of the Alps concentrated their trading activity. These representations were in the common interest of foreign merchants and the city: the merchants had a centrally located, protected place for trade and information, bargaining and contracting as well as lodging and resting; in turn, the city officials could easily survey the foreign traders, keeping an eye on whether the Venetian taxes were duly paid and

regulations respected, most of all the so-called staple right, which guaranteed the city's regional trade monopoly by requiring all wares that entered the region to be offered for sale in the city. Note that, in this sense, Venice – and notably its central Rialto district – can be regarded as one of the first places in history where to find massive "economies of scope" (in analogy to economies of scale) that are so characteristic in commerce, trade and other businesses (typically services), in which personal communication, mutual trust and effective control play a major part. Typically, these activities tend to be concentrated in relatively small urban areas or places, which, with the benefit of hindsight, can be called the birthplace of global capitalism.

2. Early on, Venice became the cradle of **banking**. Complementary to trade, tables – or literally: banks – for exchanging currencies sprouted around the Rialto. Once there, these banks had a massive incentive to facilitate trading operations by providing short-term liquidity in a relatively trustworthy traders' environment. Not surprisingly, then, the pure exchange of different monies (e.g. Venetian ducats for Florentine florins or Genoese pounds) gave birth to deposit and credit operations, to the issuing of bills of exchange and various types of short-term commercial loans. All this happened (and worked!) despite the formal ban of the Church on the charging of interest. After all, the operations could easily be designed in a way that, formally, no interest payment was forthcoming. Thus Venice became a pioneer of that typical combination of trade and short-term financial services, which – again due to massive economies of scope – tends to be concentrated in very small urban areas and in a community that provides a high level of reliable information that can be used to evaluate risks of business operations. In this sense also, Venice is a forerunner of later financial centres like, e. g., London in the 19th and New York in the 20th century.

3. Together with Genoa, Venice also became the birthplace of **long-term business finance** and, if we like, venture capital markets. That was mostly due to the nature of the maritime trade ventures. After all, these were high-risk operations that required enormous sums of capital to cover the fixed and variable costs of ships, crew, military protection and possibly also some kind of insurance. The array of new instruments of modern financial capitalism was indeed large. The types of new contracts and risk sharing ranged from various forms of sea loans up to the sophisticated "commenda", which was tailor-made to the demands of the day: a limited partnership in which one party – the one who travelled – bore the risks of labour and management on board while the other party (the one who stayed at home, but provided the capital) bore the capital risk, but was entitled to a fixed share (usually 2/3) of the profits. This spread

the risk, but kept up the incentive for the travelling “manager” to act in a way so as to maximize the likelihood of a successful voyage.

4. Apart from its commercial rise, Venice became a **centre of industrial production**. Due to its excellent trade links, which allowed to import many raw materials for industrial production from around the Mediterranean and the North at relatively low cost, Venetian manufacturing excelled in a broad diversified range of high-value added branches that included high-quality textiles (notably those made from silk), but did not exclusively rely on them as did the about equally-sized rival city of Florence. From early on, metallurgical industries and the production of chemicals – in the sense of converting one material into another – figured prominently in Venetian industrial structure, above all the making of glass in the famous works on the island of Murano vis-à-vis the city, but also the production of soap, dyes, tiles and bricks as well as saltpetre. All these crafts required special experience-based knowledge, and it is not surprising that, in these circumstances, an innovative climate developed that generated path-breaking inventions like, e. g., the spectacles, which opened the door for Venice to become a specialised manufacturer of (expensive) lenses for eyeglasses. Naturally, the city was also a major (and highly innovative) producer of ships and boats. These were built and repaired not only in its famous shipyard, the communally-owned “Arsenal”, which – after its enlargement in the 13th century – became and remained the only really large-scale industrial plant of the Middle Ages. Beside it, there were quite a few private shipyards that tended to be specialised in larger vessels, while the Arsenal produced the standard merchant galleys.

5. Parallel to the rise of crafts and manufacturing, there was a remarkable growth of **guilds**, both in number and in size. In this respect, Venice was not much different from other leading cities in northern Italy. However, due to the overwhelming political power of merchants, who had no need for associations to pursue their interests, Venetian guilds were confined to craftsmen and artisans. Historians speak of more than 100 guilds that existed in the city, with a large variety of functions and structures. Economically, the role of the guilds had a (positive) growth-conducive and a negative (protectionist) side. Common to most was a strong and positive role in the organisation of apprenticeship, which was vital to transfer knowledge across generations in a world where labour skills were mostly based on practical experience – and not on theoretical understanding that could be taught in schools. However, common to most was also a protectionist side of restrictively regulating the access to the practice of crafts and artisans. Be the balance as it may, Venice was certainly in the forefront of

this new type of “cartelized” industrial labour organisation, which can safely be interpreted as the forerunner of the labour unions that rose in the course of the industrial revolution in the 19th and the 20th century.

So much for a brief summary of the major characteristic achievements that can be identified from the economic rise of Venice. As to its great competitor Genoa, a similar success story could be told. Sure enough, it would contain many major common elements which would be tedious to repeat point by point. Basically, the economic dynamics of seaport cities was fed by the same fundamental sources of that time.

Of course, there were also some specifically Genoese contributions to economic history, which are worth to be mentioned. They mostly stem from two facts. Firstly, Genoa’s maritime realm may have been looser, but it was somewhat wider than that of Venice – stretching from the Atlantic Ocean to trading bases on the North shore of the Black sea and on the East coast of today’s Turkey (“Asia minor”) to the western Mediterranean (see again FIGURE 10). Secondly, compared to Venice, Genoa was much less stable, both in its internal politics and its external economics. Internally, the city was subject to many violent strives between leading families, which was much different to the remarkably consensual politics of the Venetian aristocracy. In this sense, Genoa resembles much more most other northern Italian cities such as Florence whose medieval history was also anything but peaceful.

However, it may have been just these two facts – the wide stretch of its maritime grip and “individualistic” flavour of the city – that were particularly conducive to remarkably adventurous careers of some Genoese citizens at the time. Leaving aside the famous Genoese seaman and discoverer Christopher Columbus, the most notorious case in terms of economic performance is the Genoese merchant Benedetto Zaccaria and his son who, in the late 13th century, set up a huge alum refinery near Phocaea in Asia minor (see FIGURE 12). Note that alum was a major input in textile manufacturing – in plenty supply on the shores of the eastern Mediterranean, but not at all in Flanders and the rest of north-western Europe where it was badly needed in the production of clothing. Over decades, Zaccaria’s refinery in Phocaea sent many heavy ship loads of alum all across the Mediterranean and along the shores of the Atlantic Ocean up to Bruges in Flanders. Apparently, the quantities were substantial: around 1330, the refinery in Phocaea produced about 700 tons of alum per year, which summed up to an aggregate value of 50.000 Genoese pounds a year. That was an enormous sum of money by any medieval standards, which must have made the Zaccaria family enormously rich. This is one of the earliest historical examples of a fortune made by a single merchant family in a combined mining, refining and trade business.

Apart from daring business ventures, the city of Genoa was also the birthplace of highly innovative new branches and techniques of business. Two of them are of particular long-standing importance:

- Genoese merchants are credited with the invention of **insurance**. Given the ever more frequent ventures in maritime trade and commerce, the range of operations became large enough to allow for some economically viable spreading of risks to be bundled into reasonable insurance schemes, and not just gambling plots. Note that this was a remarkable achievement, not least in view of the enormous risks involved in virtually any single maritime business trip. These risks were apparently so high that many maritime journeys were made in convoys with military assistance (and thus high costs of security). Nevertheless, insurance schemes developed, and where there to stay.
- Genoese (and Florentine) merchants are considered to be the inventors of **double-entry bookkeeping**. Apparently, the need for a systematic account of ever more complex commercial activity had become so desperate that some smart businessman or clerk began to give up the traditional form of unsystematic sequential memory notes and shifted to well-structured credit and debit accounts. To be sure, this was a fundamental step forward towards a rational and quantitative approach to business. The fact that it took centuries until it had spread all over Europe only shows how big a step it was.

To sum up, it is all too obvious that, from the 13th to the 15th century, the northern Italian seaports Venice and Genoa were at the top of a capitalist development that left deep traces in the further economic history until today. Together with the other leading northern Italian cities, notably the big industrial centres Florence and Milan, the two seaports were the pioneers of modern global capitalism including banking and some nascent form of insurance. Clearly, they became a kind of school for entrepreneurship whose reputation spread all over the continent. In fact, for the sons of leading merchants in the rest of Europe, some time of apprenticeship in one of the great Italian commercial cities was a highly appreciated part of education.

Naturally, economic growth and development were not costless as, with business success, there comes also the risk of failure. And it is no coincidence that the first capitalist crises are well documented for Florence in the 14th century when some major partnerships collapsed.

Be that as it may, there can be no doubt that these crises occurred when already a very high level of prosperity had been reached by the standards of the Middle Ages. After all, the medieval spurt of urbanisation and commercialisation had shot up the major northern Italian cities to a degree of specialisation, a level of productivity and a standard of living that would have been unimaginable only three centuries earlier and was still miraculous by the standards of that time in other regions of Europe.

To be sure, the success story did not last indefinitely. However, the relative decline of Venice and Genoa is largely a part of a much broader story: the relative decline of northern Italy as a whole, both as a leading region of manufacturing as of trade. We will return to this topic in section 2.3.1.2 of this manuscript when dealing with the rise of Holland and England.

2.2.4.2 Northern Europe: The Hanseatic League

In many geographical respects, there is a remarkable similarity between the northern seas of Europe, i. e. the North and the Baltic Sea viewed as one whole, and the Mediterranean (see FIGURES 12 and 13). Both are landlocked oceans, which are subdivided in two parts by a central peninsula that stretches far out: Italy in the case of the Mediterranean and today's Denmark and the very north of Germany (north of Hamburg) in the case of the northern seas. It may therefore be no coincidence that there is also a historical parallel between the rise of the northern Italian seaports and the successful association of cities under the leadership of Lübeck that we call the Hanseatic league (or, in German: just "Hanse").

To be sure, the parallel is highly seductive: Venice and Lübeck shared a similar geo-economic position with major sea links to the East, a relatively well-established network of overland trading routes leading into mainland Europe, and a somewhat more difficult, but manageable access to the western part of their maritime hemisphere. Also, the many northern and western German cities that were members of the Hanseatic league formed a network that appears to have much in common to what happened in northern Italy (FIGURE 14).

In one major economic respect, however, there was an important difference that did not disappear all over the Middle Ages and even much beyond – until, in fact, the relative economic decline of northern Italy had forcefully begun in the early 17th century: the degree of urbanisation and the standard of living. At no time did the cities of the Hanseatic league reach anything close to the per-capita income levels that prevailed in Venice, Genoa or any other of the northern Italian big cities. There was and remained a South/North-gap, and that gap can be recognised in virtually every aspect of economic development that we look at – from the

role of banking, finance and insurance down to trade volumes and business techniques. E.g., there is some statistical evidence that, in the 13th and 14th century, the value of traded goods going through the port of Genoa was in the magnitude of ten times higher than that going through the port of Lübeck, the “queen” of the Hanseatic League. And at no time of the Middle Ages are there any traces of the new technique of double-entry-bookkeeping in Hanseatic business offices. No doubt, it was northern Italy that formed the forefront of global capitalism, and not the Hanseatic league.

What then makes the Hanseatic League interesting in a lecture on Globalization, despite its relative backwardness? The answer lies in its form of organisation: it was the first highly successful voluntary association first of merchants and then of trading cities to pursue specific commercial interests. Note the difference between Venice and the Hanseatic league: Venice provides the paradigmatic case of a maritime empire, the Hanseatic league the earliest case of a voluntary trade association on equal terms, though under the leadership of some “big players” like Lübeck, Hamburg and at times Cologne.

The rise of the Hanseatic League must be seen against the background of the weak central power in the Holy Roman Empire of German Nations, which was unable to provide military security and protection against pirates and other dangers on long-distance maritime trade routes. From the 12th century on, German merchants (at the time called: Saxons) began to compete with the Frisians and Vikings, who so far dominated the trade in the northern seas. Apparently, there was a strong need for mutual support and protection. Thus, in the mid 13th century, a first German merchant co-operative was founded on the island of Gotland (near mainland Sweden). This happened in a period of massive German commercial expansion. At about the same time, merchants from Cologne received privileges by the King of England for trade in London, and the town of Lübeck was founded as a deliberate attempt of the duke of Holstein to attract merchants and settlers from crowded Rhineland and Westphalia to the shores of the Baltic Sea.

Gradually, a long-distance trade network emerged and grew. And with it, the merchants’ co-operative gained clout and could ever more successfully bargain for privileges. Historians tell us that, in negotiations at the English court, the German merchants came to be known under the name “Hansa” - meaning, in old Germanic language, “Schar”, i. e. in English something like crowd, horde or – in a maritime spirit – convoy. For more than two hundred years, the “Hansa” remained a merchants’ association before, at the so-called first official convention of city representatives in Lübeck in 1356, it was turned into a Hanseatic League of cities, which survived for more than 300 years until it finally withered away in the second half of the 17th

century. Much later, the German word “Hanse” or “Hansa” had a remarkable comeback and re-emerged with a strong sense of pride, e. g., in a number of titles of cities and in the name of the national airline “Lufthansa”. Nowadays, there are few words of German history that carry such a positive image as the term “Hanse” or “Hansa”.

As an association of about 70 cities under the – never questioned – authority of the city of Lübeck, the Hanseatic League had remarkably few formal rules. It worked like a pragmatic cartel-like cooperation to pursue all kinds of common trade interests, from bargaining for privileges and providing protection to fighting wars and securing arbitration among quarrelling members. Most importantly, it entertained common trading bases, which were divided into two categories: a minor trading post called a “Faktorei”, and a major one called a “Kontor” (if you like: a main office). Of the major ones, there were four, and all four played a most prominent part in the history and the economics of the League: Bruges in Flanders, London in England, Bergen in Norway and Novgorod in today’s Russia (see FIGURE 15). Note that the maintenance of these four main offices involved substantial direct foreign investment. Quite literally, each of them consisted of one or more buildings and a yard, which were kept up by the Hanseatic League over a long time and became duly famous: the “House of the Osterlinge” in Bruges – “Osterlinge” being the name that the Flemish gave the Hanseatic Germans –, the “Steel Yard” in London, the “German Bridge” in Bergen and the “Peter Yard” in Novgorod.

To understand the structure of the trade that was carried out by the Hanseatic League, it is first of all important to recognize the strategic role of each of these four major trading posts. Bruges, which was clearly the most important of the four, opened the commercial gate to the products of the textile industry in Flanders, which were to be sold everywhere else. As a large city in a densely populated and urbanised area, it also provided the biggest market for basically all the products that came from the other major trading posts (and elsewhere): high-quality wool from English sheep, which was bought in London and sold in Bruges as a major input for the textile industry; ample fish supplies, notably of stockfish, which were bought in Bergen; and grain, furs, pitch, tar and wax, which were bought in Novgorod.

Of course, this was not all that happened in trade. Important commercial links existed between the Hanseatic cities themselves: early on, a most profitable salt trade from the hinterland of Lübeck (notably from Lüneburg) and from there to whatever destination; imports of grain and timber from the rural hinterland of the Hanseatic ports of Stralsund, Stettin, Danzig, Königsberg, Riga and Reval on the Baltic Sea, whose eastern coast also delivered the supplies of amber that were exported all over Europe. In later years, there were massive imports

of salt from Bourgneuf and Brouage at the Atlantic coast of France, which were shipped to Bruges and – via Lübeck – to eastern markets. And there were large purchases of high-quality salted herring, which was traded in southern Sweden at the regular fairs of Falsterbo and Skanör and sold all over central, eastern and western Europe.

From about 1200 to 1500, which are the centuries of its most remarkable commercial success, the geographical extent of Hanseatic trade widened continuously. FIGURE 16 shows how this proceeded. Early on, everything was focused on the big northern maritime axis London-Bruges to Reval-Novgorod, which contained only a very short land passage between Hamburg and Lübeck, which in later years was increasingly replaced by the Skagerak-route around Denmark. Gradually, the League began to establish major overland trading routes, both in western and central Germany: the straight North/South-route from Lübeck and Hamburg via Magdeburg, Erfurt and Nuremberg to Venice; and the North-West/South passage from Bruges over Cologne, Frankfurt and Milano to Genoa. In eastern Europe, the League created the routes from Danzig to Breslau, Cracow, and Lemberg, and from Riga to Smolensk. Clearly, at a time of very little central authority and protection, the firm establishment of all these trade links were major economic achievements.

Despite its apparent success, the Hanseatic League did gradually lose its economic power and rationale. This was so for essentially two reasons, a political and an economic one:

- Politically, the power balance of cities on the one side and territorial rulers on the other gradually shifted in favour of the latter. This meant that it became increasingly difficult for Hanseatic cities to provide credible political – and possibly military – support to each other. Quite often, they found themselves put under heavy political pressure by their territorial rulers, who had other political priorities than urban trade and commerce. In the end, most of the Hanseatic cities were simply “swallowed” by larger territories that provided their own “public goods” of law and security and pursued their own policies, which were less commerce-friendly. Note that, very much like Venice, Hanseatic towns were typically governed by merchant families who had strong interests in putting commercial success on top of all political goals. Clearly, there was much less point for later territorial rulers of large countries to keep up close commercial ties to far-away towns in possibly hostile lands.
- Beginning in the 16th century, the poles of economic growth in Europe began to move westwards – to Holland and England, and away from Italy and Flanders, which played such a strategic role in the Hanseatic commercial network. After all, the Hanseatic

trade was mostly transit trade of a very specific kind: with a bit of simplification, one may say that it was trade in the sense of the Heckscher/Olin-theory, i. e. labour-intensive manufactures (produced in Flanders) in exchange for raw materials and agricultural produce. However, once the dynamics of manufacturing moved to different places where local commercial centres stood ready to grow into carrying out the complementary trade (think of Amsterdam and London), then the monopolistic margin of Hanseatic expertise was to melt away. In this sense, the decline of the Hanseatic League was part of a broad shift of regional growth patterns in Europe to which we shall return below in section 2.3 of this manuscript.

2.2.5 Land Trade

From today's hindsight perspective, it is very hard to imagine the sheer extent of the difficulties that travelling overland incurred in the Middle Ages. After all, nothing comparable to the ancient Roman road network existed anymore, and not before the 19th century were roads between cities systematically kept up for commercial use. In some places and some times, there were Royal rules that required from local landlords some minimal road maintenance; but - even if they existed - the rules were not properly obeyed, and the central authority had neither the means nor the power to enforce them.

In these circumstances, most roads were no more than loose tracking paths that crossed the countryside. Whether they could be used depended on a combination of conditions that were determined by the season, the weather and the nature of the soil. In Flanders, e. g., the heavy clay soil combined with the humid climate made it virtually impossible to carry food and merchandise overland in winter time. Thus the merchandise was reloaded onto boats and barges wherever a river allowed some sort of shipping. As a consequence, commercial traffic took place even on very small streams and brooks, which today would be regarded as completely inadequate for navigation purposes. However, even on big rivers there were major obstacles and barriers to movement and trade. Occasionally, natural variations of the level and speed of water flows made downstream navigation dangerous and upstream navigation impossible. And local landlords typically installed toll and staple stations at which a tariff had to be paid and/or the merchandize offered for sale.

Despite all these hindrances, well-defined patterns of long-distance routes overland combining roads and rivers gradually emerged. These patterns were well-known and occasionally summarized in handbooks that gave travel advice to merchants. E. g., the Itinerary of Bruges (FIGURE 17) shows major trading and travelling routes from Bruges to towns and cities all

over Europe. As such, it nicely demonstrates that, even for very remote places like Seville, Constantinople and Moscow, there were quite precise conceptions on how to get there.

While the technology of maritime navigation made remarkable progress in the later Middle Ages and in Renaissance times (see section 2.2.2.1 above), the same cannot be said for the technology of overland travel. Nevertheless, it would be misleading to conclude that travel costs remained altogether unchanged on the main trans-European overland routes. To be sure, the increasing traffic alone was likely to lead to some cost reductions in terms of an improved knowledge of local conditions and better service stations along the roads as well as some efforts to widen the worst bottlenecks. E. g., in the mid 13th century, travelling between central Germany and Italy was made much easier as the St. Gotthard pass was added to the supply of trans-Alpine routes, and that was only possible because heavy investments in road and bridge construction at the rugged northern approaches to the pass had been made beforehand. Apparently, local authorities were not blind to the needs of mounting commercial needs – and to the profits that could be made from providing at least a minimal road infrastructure and charging the travellers for its use at appropriate toll stations.

2.3 Towards the Modern Age (1500-1820)

In the economic history of Europe, the three centuries between Renaissance times and the industrial revolution are best viewed as a transition period. To be sure, contemporaneous observers could not see them in this way because they did not yet have the experience of the altogether dramatic changes that took place from the 19th century onwards. With the benefit of hindsight, however, the major elements of transition become clearly recognisable. They are to be found in three fields:

1. Demographically, not much changed for a long time. Sure enough, the population size of Europe recovered from the Black Death. Beyond that, there was a slight upward trend in the 16th century. Then again came a temporary stagnation, and not before the 18th century did a general forceful upward trend set in, which eventually accelerated to lead into the sharp increase that was to be observed in the 19th and most of the 20th century.
2. Technologically, there were gradual improvements made in many fields, but they did not add up to anything revolutionary in terms of productivity growth and industrial organisation. A merchant, craftsman, manufacturer or peasant - if miraculously trans-

planted from Europe in 1500 to Europe in 1750 - would not have been utterly disoriented in the new world: a few modern techniques to get acquainted with, but that would have been a manageable task for him. After the mid 18th century, however, the speed of change increased markedly and further accelerated in a genuine technological revolution that followed in later decades.

3. Economically, the conditions did not change much, again until roughly the mid 18th century. Of course, there were regions that rose and others that declined for reasons that we shall analyze further below. But, all in all, the average standard of living in 1500 had not been all that far below what it was in 1750. From then on, however, it began to rise, and that rise continued at a much accelerated rate with the so-called industrial revolution, which gained breath and depth in the first decades of the 19th century and beyond.

Of course, these roughly parallel developments of demography, technology and economics point to some common causes, which will be a major subject of section 3.1 of this manuscript when we look at the aftermath of the so-called industrial revolution. As we shall see then, however, these causes did not fall from heaven, but were the consequence of developments that are rooted in earlier times. Hence, with the benefit of hindsight, the three centuries between Renaissance times and the industrial revolution become something like a period of "limbering up" to the challenges that later history had on offer:

- vast new territories were discovered overseas and to some extent integrated into the trade networks, but that did not yet really lead to a genuine Globalization;
- great scientific advances were made and an enlightened intellectual climate developed but that did not quite lead to major breakthroughs in technology,
- the limits of growth in terms of an energy crisis were touched, but that did not yet lead to a broad-scale switch to inanimate power that was to follow later.

To be sure, a period of transition is almost as fascinating to study as a period of breakthrough and achievement, not least because it is typically in periods of transition that major switches are set, which have very important consequences for future developments. After all, history is "path-dependent", i. e. it contains many irreversible traces that pre-determine the course of events. In this sense, one major consequence of the period under consideration is the relative decline of the Mediterranean with Italy as its focal point of prosperity, and the rise

of what may be called the Atlantic economy with first Holland and then Britain and then many others like the United States and Germany in the lead. In fact, the reversed economic weight between “North” and “South” in Europe is still a trait of recent times, and not before the very recent rise of Spain, Portugal and Greece within the European Union did it lose something of its sharply dividing character.

2.3.1 Europe’s Overseas Expansion

When speaking about the history of Globalization, one may be tempted to regard the period of the great discoveries in the 16th and 17th centuries as an absolutely crucial one. After all, historians agree that, by that time, the dynamic part of Europe had definitely overtaken the more ancient, but economically stagnating civilisations of China and Byzantium (by then the Ottoman empire) in terms of prosperity and living standards. And once this dynamic Europe was to discover and conquer big chunks of the rest of the world and embark on colonial ventures, this might well be regarded as the birth of Globalization proper.

In a geographical and political sense, this may be true. Economically, however, things are more complicated. As we have seen, important networks of long-distance trade integration had emerged and grown much earlier, notably in the Middle Ages. In turn, a breakthrough to really modern forms of Globalization would have required a new stage of industrialization and of transport and communication techniques, which was clearly not yet on the horizon in the 16th and 17th centuries. This is why Europe’s expansion overseas at that time has much more the character of a “follow up” to the Medieval and Renaissance period than of a qualitatively new stage of economic history.

Be that as it may, Europe’s overseas expansion is anyway important for other reasons. To see this, we have to distinguish between two types of expansion, which were very different in character and consequences: that of south-western Europe (Portugal and Spain) and that of north-western Europe (Holland and England).

2.3.1.1 Portugal and Spain

Historians agree that the extremely successful ventures by Portuguese and Spanish conquerors over the Atlantic had one major technical and military reason: the development of a robust sailing ship that could carry guns and cross oceans. Apparently, this was decisive for the crushing of native resistance and the founding of vast colonial empires in later-called Latin America (FIGURE 18). And it ushered in a genuine boom period on the Iberian peninsula

that lasted until the early 17th century. The boom was mostly fuelled by a massive inflow of gold and silver: in America, rich mines could be exploited, and the precious metals sent by the shipload to Europe. Once there, the gold and silver was mostly minted into coins, and these were issued as fresh money.

The import of precious metals by Spain alone was massive (see FIGURE 19). And so were the macroeconomic effects of the bullion bonanza. With fresh gold and silver money, the Spanish and Portuguese could buy goods everywhere in Europe and beyond. In macroeconomic terms, the Iberian peninsula ran a large current account deficit; thus the money was spread all over Europe and beyond, and the boom fuelled production everywhere. However, with population only moderately increasing and productivity growth proceeding at the very slow pace of that time, the boom was bound to lead to a persistent inflationary pressure. In fact, this happened and gave an economic name to the 16th century: the age of the “price revolution”.

By the inflationary standards of the 20th century, the term “price revolution” is an exaggeration. However, by the standards of earlier times, it is fully justified: between 1500 and 1620, prices rose by 300-400 % in Europe, which amounts to an annual average inflation rate of 1.2-1.4 %. Note also that the ample supply of liquidity led to very low interest rates. A study of the financial market of Genoa for this time has shown that, in the early 17th century, nominal interest and discount rates had come down to a low of about 1.5 %, probably lower than ever before. This shows that the real interest rate (i. e., the nominal interest rate minus the rate of price inflation) must have come close to zero. Interestingly, this indicates that price inflation was not yet strong and sustained enough to be anticipated in inflationary expectations of lenders, which would otherwise have demanded higher nominal interest rates as typically happens in highly inflationary regimes (the so-called Fisher effect).

Be that as it may, the macroeconomic boom all over Europe did hardly bring long-term benefits for those countries that fuelled it by their transatlantic conquests. In fact, the failure of Spain and Portugal to profit in terms of long-term growth from the sudden richness is a classical textbook example of how awfully a macroeconomic boom can fail to develop a country. Apparently, the productive capacity of the Iberian peninsula did not at all react elastically to the welcome tide of demand: skilled workers were lacking, and there was no discernible political strategy by the rulers to seize the opportunity to increase the supply of human capital, e. g., by educating people or by attracting foreign craftsmen, artisans and entrepreneurs. To the contrary, there is much anecdotal evidence that the Spanish nobility regarded imports of

fancy goods as something to be proud of, and not as a potential loss of opportunities for development.

By the mid 17th century, the inflow of precious metals had turned into a trickle since the major mines were depleted. Quite abruptly, the party was over. For the Iberian peninsula, the result was depressing: while other parts of Europe had thankfully accepted the additional demand for their products and used it to push forward their own economic development, Spain and Portugal had not advanced their productive potential at all. It is remarkable that it took about 350 years until, in the last quarter of the 20th century, the two countries did eventually embark on a successful growth spurt that is belatedly making them catch up with the rest of western Europe.

2.3.1.2 Holland and England

The story of Holland is the exact reverse of that of Portugal and Spain. Long before the colonial period, the Dutch economy was relatively advanced in terms of both productivity and trade integration, though somewhat behind its highly urbanized southern neighbour Flanders. In the second half of the 16th century, political circumstances tilted the balance in favour of the north. The rebellion of the low countries, i.e. of both the northern part "Holland" and the southern part "Flanders", against Spanish rule led to much damage in the south, i. a. the sacking of Antwerp, which had overtaken Bruges in importance since the decline of the Hanseatic league. Still more importantly, the South remained under Spanish rule, while the North became independent as the United Provinces, which – being predominantly protestant – practiced a very high degree of religious tolerance. As a consequence, a wave of highly qualified artisans and craftsmen moved from the south to the north, thus tapping the enormous "human capital" of highly developed Flanders for the benefit of the United Provinces.

Parallel to these fortunate circumstances, long-standing indigenous advantages now showed their full potential: highly productive agriculture and fishing and an already well-developed industry, which ranged from food processing and grain milling up to the manufacturing of textiles and clothing. A major asset for industry was the cheap and ample supply of energy: historians tell us that, by 1630, there were 222 industrial windmills in use in Holland (not including those mills that were applied to grind grain and to provide drainage). On top of all this came a pronounced maritime orientation, with a strong tradition of seamanship and shipbuilding. In fact, the Dutch had developed a worldwide reputation as leading cartographers and specialists for geometrical and astronomical instruments.

In shipbuilding, the Dutch were responsible for a major innovation of the 17th century: the so-called fluitship (in Dutch: *fluyt*). In the Hanseatic League, the main cargo carrier had been the cog (in German: “*Kogge*”), which was a wide-bodied sailing ship with one mast. By medieval technical standards, it delivered an optimal trade-off between gain in loading capacity and loss in speed. In many respects, the fluitship followed the same philosophy, but much improved the trade-off by enlarging and optimizing the ship’s body and adding a second mast, which was standard practice in the 17th century. Compared to its size, the ship was extremely light and easy to handle so that the costs for the crew could be minimised. On the other hand, it was slow and defenceless but that was not so much of a risk because the Dutch navy provided protection on the main routes of commercial traffic.

Given their rich heritage of industrial and maritime experience, the Dutch were well equipped to take full advantage of their colonial expansion, which led them notably into South-East Asia (see FIGURE 20). More than that: the Dutch were able to develop an unprecedented expertise in world-wide commodity markets. Economic historians agree that Amsterdam in the first half of the 17th century is a paradigmatic example of what may be called early global capitalism – with professional trading facilities that were much more advanced than anywhere else in the world including the Italian seaports Venice and Genoa. In fact, the port of Amsterdam became a major international market for a vast variety of goods from different countries: Swedish and Japanese copper, Baltic grain, Italian silk, French wines, Chinese porcelain, Brazilian coffee, oriental tea, Indonesian spices, Mexican silver etc. From 1585 onwards, the Amsterdam Commodity Exchange published bulletins of prices, which were read all over Europe and contained – by 1634 – no less than 359 commodities.

To be sure, Amsterdam was the main hub for the new product world that was opened by colonisation; and as such it became a school for entrepreneurship in everything that was needed in worldwide commerce and finance. For the Dutch economy as a whole, however, the more down-to-earth link between trade and the large rest of the economy was possibly of even greater importance. In fact, Holland did not only have skilled merchants, but also a manufacturing sector that was able to take full advantage of an integration into the world economy. Historians agree that Dutch manufacturers were the first to rigorously apply a philosophy of mass production – long before the industrialization allowed a maximal exploitation of technical economies of scale.

With some courageous simplification, one may say that the Dutch were the first traders in the world to put their absolute priority on cost-cutting to sell at low prices, even if the quality of their products was still to some extent inferior. Note that this was an altogether new approach

to world markets. Traditionally, the costs of transport were high enough to make such a strategy appear to be not very promising. Nowadays, the perspective has completely changed in view of low transport and communication costs so that we take a low-cost strategy for granted – as a natural and rational option for producers in laggard countries to prise themselves into markets. Historically, however, this was not so, and the Dutch strategy had therefore much innovative content in the 17th century. For the first time, long-distance commerce really lost its image of luxury, which had accompanied world-wide trade since early medieval or even ancient times.

Of course, it is difficult to speculate about the reasons for the typically Dutch approach. On the supply side, the low cost of energy due to the abundant supply of powerful wind mills may have given the Dutch a head start in price competition and some scope for realising economies of scale. On the demand side, the relatively egalitarian distribution of income in a society that was dominated by many well-to-do merchants, peasants and craftsmen rather than a few landed aristocrats may have tilted the balance towards the needs of an emerging middle class rather than elitist tastes.

Whatever the answers to these difficult questions, there were strong forces in the dynamics of the world market that worked in favour of the Dutch strategy. To see this, one has to take a side look at the decline of northern Italy, notably in the manufacturing of textiles, the most important of all industries at that time. To be sure, Italy was and remained a high-quality producer of textiles and clothing, but as such, it was unable to compete in an ever more integrated world market for textiles that was increasingly dominated by Dutch (and also English and French) producers. Not only did the volumes of Italian textile exports decline, but the markets in Italy itself were increasingly flooded with cheaper products from north-western Europe. In fact, the statistics of cloth production in Italy itself indicate a really disastrous decline in the 17th century (see FIGURE 21). In a sense, northern Italy became de-industrialised, and the trade structure of the region strongly tilted towards raw materials as inputs for others' manufacturing (like raw silk) as well as agricultural produce like oil and wine.

As a matter of fact, the northern Italian cities (and to some extent also the Flemish ones) in the 17th century deliver the first major example in industrial history of a highly developed region with a proud manufacturing tradition that was not up to the economic challenges of its time. Historians see mostly internal economic reasons for this failure: an excessive control of guilds, which had become a major obstacle to technological and organisational innovation because they concentrated on defending their narrow group-egoistic interest, a high level of

labour costs, which stifled competitiveness, and a high level of taxation, which discouraged investment. In general, it is the typical picture of a mature economy that slides into a state of “sclerosis”, with vested interests of traditional pressure groups preventing a forward-looking adjustment to the economic challenges.

Note that, in these circumstances, international trade may in fact lead to an acceleration of an industrial decline as it opens the country to import competition and thus leads to a reallocation of resources away from industry towards primary activities with much less growth potential. In a sense, the forces that led to the country's rise then also lead to its decline. In normative terms, however, this can hardly be held against trade itself. It is rather a reflection of the fact that the internal forces of the economy are not innovative enough to develop new branches of economic activity that, for the time being, guarantee a monopolistic margin to be reaped and a new resource and growth potential to be tapped.

We shall repeatedly come back to this type of situation of countries in the course of world economic development in later centuries. In most of these cases from the 19th century to the present, however, the worst outcome of such-like crises has been a relative decline of a country, with its economy still growing, but at a markedly slower rate than those of other countries. For Italy in the 17th century, however, the main indicators point to a decline not only in relative, but also in absolute terms, if only because there was not yet a strong worldwide growth dynamics that could pull forward even the laggards. That kind of general dynamics was in fact not to come before the industrial revolution.

Parallel to the decline of Italy and the rise of Holland, one may speak of a dawn of England. As in the case of Holland, the colonial expansion was only part of the story. To be sure, it took place (see FIGURE 20); and it helped notably the city of London to embark on its rise as one of the major hubs of world trade and finance, which reached its peak in the 19th century. For England as a whole, however, it was the linkage between trade and manufacturing that was of utmost importance. In this respect – and even more so than in Holland –, the textile industry played a key role.

In the later Middle Ages, England had been a major exporter of high-quality raw wool to the continent, notably to Flanders. The trade-link from London to Bruges was in the hands of the Hanseatic League, and English wool was the major input in the Flemish textile industry. In Renaissance times, this pattern began to change (see FIGURE 22). Gradually, England grew into the manufacturing of textiles and became an exporter of fabricated cloth that was then traded in Antwerp, which was replacing Bruges as the big commercial centre of the low coun-

tries until Amsterdam took over this role. Note that, in this period, the trade was not anymore dominated by the Hanseatic League, but by southern German merchants like the Fuggers and Welser from Augsburg, who had made fortunes in mining ventures in central Europe and traded in textiles and metals all over Europe. Gradually, they also began to satisfy their commercial demand for cloth with English products bought in Antwerp instead of Italian textiles that were on offer south of the Alps.

Historians agree that, in many respects, the development of the English economy and society ran parallel to that of Holland:

- an increasingly diversified economy with various growing branches of manufacturing from textiles and clothing to metallurgy, arms-making and shipbuilding;
- a strong dose of immigration of highly-skilled labour, notably of protestant Walloons and French Huguenots who were welcome in England as craftsmen and artisans;
- and a general receptiveness vis-à-vis new ideas from abroad, which is reflected in a remarkably open intellectual discourse and exchange.

All this indicates that, at least for this historical period, the focus on world-wide commerce and on colonial expansion is much better interpreted as a part of a wider modern outlook that contrasts sharply with what happened in the Mediterranean, be it the “colonial bonanza without development” of Portugal and Spain or the conservative industrial sclerosis of northern Italy and to some extent Flanders. In this sense, it was not per se the overseas expansion that ushered in a new age, but a modern rational, open and tolerant philosophy and attitude that allowed opportunities for growth to be taken wherever they emerged.

2.3.2 Scientific Advances

There were major scientific advances in the roughly three centuries under consideration. More than that: for the first time in history, something like systematic research based on rigorous theoretical reasoning and empirical evidence began to take shape. Clearly, this was a harbinger of modern science and technology, quite a big step forward from the mere fascination with mechanics of the later Middle Ages and Renaissance times, which was still moulded in a trial-and-error approach to the real world.

It is difficult to pick whatever dates, events or trends to verify this change of outlook, which had more the character of a deep undercurrent that gradually shifted the basis on which all economic, social and political activities were standing. After all, the change was subtle, but it took place in many different fields that go well beyond the scope of this lecture. Neverthe-

less, three distinctive developments are worth mentioning, if only because they are absolutely uncontroversial as major elements in the process and as decisive stepping stones to economic modernity.

Firstly, with the production of the famous Gutenberg Bible of 1452-4, printing with movable characters was invented. The new technology spread fast all over Europe (see FIGURE 23). It laid the ground for a long series of further improvements in printing technology, which led to a drastic reduction of the cost of knowledge diffusion and education. Gradually, books – and also pamphlets and papers – became available on a broader scale. Step by step, they lost their image of being something exclusively confined to the precious libraries of a super-privileged class. Although the age of full information of broad masses of people was still far away, knowledge could henceforth disseminate well beyond the limits and horizons of those islands of intellectual discourse that had developed over the centuries (e. g. monasteries). Note that, in the Middle Ages, these islands themselves had been well interconnected over long distances, not least because they used a common language, namely Latin, which was the lingua franca of the few highly educated people of the time, notably the monks. However, due to the scarcity and high price of books, this discourse remained a matter of a tiny élite and had relatively little feedback on the population at large. It is exactly this that began to change with the dawn of modern times. And the new trend was strongly reinforced by the rise of Protestantism whose emphasis on texts and books led to a new positive appreciation of literacy – and the educational efforts to achieve it.

Secondly, new tools of thought were developed, and the ground was laid for exact scientific reasoning. Mathematics, physics and other sciences made enormous progress and provided the intellectual pillars on which any precise and quantitative approach to reality was to be based. Apparently, these modern disciplines were increasingly attractive for great brains: Nikolaus Copernicus, Galileo Galilei, René Descartes, Isaac Newton, Gottfried Wilhelm von Leibniz and also Otto von Guericke form the tip of an iceberg of many more thinkers who left behind the realm of traditional dogmas and founded the world of rationally-based empiricism or, for that matter, empirically-oriented rationalism. And quite a few of them staunchly defended their modern approach against backward-oriented conservatives, who fought tough battles against the modernists. It is remarkable that the “modern” approach was not confined to high-brow philosophy and natural sciences. It also emerged in more mundane fields. E. g., the systematic recording and interpreting of data – i. e. modern statistics – begins to develop at that time, with one of the pioneers being Gregory King who developed a national accounting for England in 1688, the earliest calculation of gross domestic product of modern ages,

which we mentioned above in section 2.1.2 as our point of reference when analyzing the standard of living in the Roman empire at the time of the emperor Augustus.

Thirdly, there was a general reappraisal of the importance of crafts and technology. In the Middle Ages – and even more so in Renaissance times – there had been a strict separation between science as part of philosophy on one side and technology as part of arts on the other. In line with the tradition of classical antiquity, “philosophers” and “artisans” were considered to come from two different worlds, which were completely incompatible and delineated two distinct professional camps that had nothing to do with each other. From the 17th century on, this attitude began to change and make place for a much more respectful consideration of the work of craftsmen. Gradually, the intrinsic link between science as such and technology as applied science was recognised. This was an extremely important intellectual step forward. In the long run, it opened the gate to a “scientific” approach to manufacturing, and thus to a system of apprenticeship and schooling that did not only rely on the transfer of knowledge from experience and learning-by-doing, but also of theoretical understanding and formal analysis of engineering. Of course, it was still a long way to arrive at a mature stage of science-based-technical learning that was developed in the later stages of industrialization in the course of the 19th century. Still then, the intellectual seeds were sown much earlier in a new attitude vis-à-vis technology.

2.3.3 An Energy Crisis

For centuries and all over Europe, a main source of energy for industrial production and heating had been charcoal. In some areas like first Italy and then England, whose natural endowment with forests were quite limited anyway, this led to an ever more acute scarcity of timber. Between the early 16th and the late 17th century, the price of charcoal in England rose by a factor of about 10 (see FIGURE 24). This ushered in a genuine “energy crisis”. By that time, England had become not only an industrial centre for textiles and clothing, but also a major manufacturer and exporter of metallurgical products such as cast-iron artillery. Due to the high energy intensity of its production, the metallurgical industry was hit most badly by the surge of energy prices.

In two major respects, England’s reaction to the crisis was highly innovative. Firstly, imports of timber from (forest-rich) Scandinavia were stepped up. In a sense, Sweden became a kind of energy-backyard for Britain, which made economic sense due to the relatively low cost of maritime transport of bulky commodities over the Baltic Sea and the North Sea.

Secondly, England began to substitute coal for charcoal as a major supply of energy. To be sure, coal had always been in ample supply in Britain as there were massive coal fields notably in northern England along the river Tyne (See FIGURE 25). For a long time, the existence of these supplies had been well-known in the country, but the motivation to burn coal on a broad scale was nevertheless very limited. The reason for this sounds almost modern: there was a widespread belief shared all over Europe that the fumes from coal-burning were highly poisonous, which - though not quite true – echoes more rational environmental and health concerns of modern times. In view of the sharply rising costs of charcoal, however, the necessity of using coal was becoming ever more pressing so that the suspected hazards of coal-mining were ever less regarded as a prohibitive hindrance.

From the 16th century onwards, England became the first European country to extensively rely on coal as a major source of energy supply. Between 1550 and 1630, annual coal production in Britain rose from about 210,000 to 1,5 million tons, which comes down to an annual growth rate of 2.5 percent – an enormously fast growth at a time when annual economic expansion for the economy as a whole was still negligibly small (though in general positive). At a similar speed, maritime coal shipping from Newcastle to London increased from about 35,000 tons in the middle of the 16th century to about 560,000 tons by 1680, which amounts to an annual growth rate of 2.2 percent.

With the benefit of hindsight, historians agree that this structural change on the energy supply side – the rise of “King Coal” – had vast consequences for the further economic developments all over the world. In the long run, it pushed the door open for a massive rise of manufacturing productivity, and it laid the ground for large-scale industrialization, which started on the British isles. Also, it fundamentally changed the economic geography of the country. While coal could be shipped from the port of Newcastle to the port of London at relatively low cost due to the advantages of seaborne trade, transporting coal in general – and notably overland – was hugely expensive. As a consequence, energy-intensive production lines were bound to move to those places where ample supplies of coal were available or could at least be made available at low cost due to easy water transport on rivers nearby. Hence, much more than anything else, the supply of coal was to determine the newly emerging patterns of economic geography and urbanisation until well into the 20th century. We shall return to this matter further below in section 3.1 of this manuscript.

It is important to realise that Britain’s turn to coal is a major example in economic history for a successful overcoming of a serious energy crisis and the concomitant “limits to growth”. From today’s hindsight perspective of modern ecology, it may be tempting to regard it as an

unfortunate detour via a “dirty” technology, which until nowadays has led to a long sequence of environmental problems, occasional disasters and maybe even global warming. However, such an interpretation would be economically and historically misplaced. It would miss the fact that – besides water, wind and wood – there were no other sources of inanimate energy available than coal. Sure enough, water, wind and wood had their obvious limits, as they incidentally still have today after centuries of further remarkable technological advances. Without tapping the coal reserves, a development like the industrial revolution would simply not have come about, and neither would have all later shifts to whatever cleaner technologies. Thus interpreted, Britain’s turn to coal was not only necessary, but highly innovative: in a path-dependent world, where technical and economic options emerge in a way that is essentially unpredictable, British society proved open enough to overcome traditional prejudices and venture into a completely new world of industrial organisation and growth. Even with the benefit of hindsight, it is very hard to see any promising alternative for the English economy at that time.

2.3.4 The Infancy of Modern Industry

Of course, the transition from a pre-industrial to an industrial society and economy was gradual and not abrupt. And neither was the transition to a much higher degree of international economic integration, which for the first time in history may duly deserve to be called “Globalization”. Notably in England, but also in the most advanced regions of continental Europe, there were signs of a fundamental change of technology and economic activity by the mid 18th century. From then on, the indicators multiply and cover an ever wider range of economic and social phenomena. With a view to Britain up to the period of the Napoleonic wars, three points stand out in this respect:

- From about 1770 onwards, major technological innovations begin to change the management and organisation of some important branches of industry. In textile manufacturing, inventor-entrepreneurs turn the spinning of cotton yarn into one of the first industrial activities to be mechanised through extensive use of machinery (first Richard Arkwright’s “water frame” and then Samuel Crompton’s “spinning mule”). Henry Cort, an iron and steel producer, invents the “puddling technique”, which allows to transform pig iron into steel much more effectively and at a better quality than was previously possible. James Watt and other mechanical engineers first develop and then improve the steam engine, which finds its first major use in driving water pumps in tin, copper and coal mines.

- At about the same time, steam power begins to be used for transport purposes. First “street-cars” and “locomotives” are constructed and put to use, but they fail to reach technical and economic maturation not least because there are no proper roads or tracks to use them. On water, progress is slightly faster: in the United States, the first paddle-steamship goes into operation on the Hudson River in 1807. By and large, however, the attempts to use the new source of inanimate power for mobility and transport are still highly experimental and do not yield results that are comparable in their revolutionary extent to what is happening at the same time in some manufacturing industries. Nevertheless, they are forerunners of what was to come a generation later.
- Intellectually and politically, a new consciousness develops with respect to the operation of a market economy in an industrialising framework. Quite literally, the modern science of economics is borne with the famous book “The Wealth of Nations”, published in 1776 by the Scottish moral philosopher Adam Smith. The book delivers a first account of an optimal division of labour through trade, commerce and specialised production that comes about through the working of self-interests in a market economy. What is more, the book is widely discussed and positively received in political circles and in British society at large. At about the same time, the British parliament passes laws against industrial espionage, a clear sign of the increasing awareness of the importance of technological knowledge for international competitiveness. Despite these laws (and a long-standing framework of patent protection that went back to 18th century legislation), the spreading of knowledge turns out to be hardly controllable, and so does the spreading of industrialization.

Despite these early harbingers of a new age, it is justified to take the time immediately after the Napoleonic wars as the real beginning of that combination of industrialization, integration and Globalization that stands at the centre of our interest in this lecture. From that time on, a straight line of reasoning leads up to the present.

3 Globalization I (1820-1990)

3.1 The First Era of Integration: 1820-1914

Politically, the almost 100 years between the end of the Napoleonic wars and the beginning of World War I were a time of relative peace in Europe. No major international conflicts hap-

pened, no major intra-national revolutions occurred, and not so many major changes of political geography can be recognised when juxtaposing the map of the continent after the Vienna Congress in 1815 and on the eve of World War I in 1914. True enough, new nation-states were founded, the biggest ones being Italy and Germany. And the waves of national interests and sentiments behind the attempts at nation-state building brought about numerous minor wars, skirmishes, threats and diplomatic crises, which eventually led down the road to World War I, the most destructive military conflict the world had hitherto seen. Up to that point, however, things were running rather peacefully. By the political standards of earlier and later times, there can be no doubt that the period was remarkably quiet.

Economically, things look completely different. With the benefit of hindsight, we clearly recognise in these nearly 100 years the first time of a dynamic growth of Europe and some parts of the world. By the term “dynamic”, we mean an improvement of economic performance at a speed that is

- much higher than what prevailed before (although, alas, we have anything but reliable numbers on earlier periods!), and
- roughly comparable to what we have become used to in the industrialised part of the world in more recent times.

What are the magnitudes of growth rates we speak of? An idea of them can be gained by looking at the data on national accounts that have been carefully collected, partly even literally constructed by the British economic historian Angus Maddison (whom we may respectfully remember from section 2.1.2 as that specialist for quantification who went as far as estimating a measure of gross domestic product for the ancient Roman empire!). In a millennial volume published by the OECD (Organisation of Economic Cooperation and Development) after 2000, Maddison presented his own calculations of annual rates of growth for the relevant periods under consideration. This allows us to gain a rough quantitative picture of how the different periods under consideration do compare in terms of economic performance.

Two of Maddison’s major tables are presented in our FIGURES 26 and 27, which show the rates of growth of GDP (gross domestic product) and GDP per capita for a number of European and non-European countries in the periods 1500-1820, 1820-70, 1870-1913, 1913-50, 1950-73 and 1973-98. For western Europe excluding the Iberian countries (in the tables called: “12 Countries Total”), the message of these statistics comes down to two points:

- The period 1820-1913 ushered in a new era of growth. Real GDP grew much faster than before – at average annual compound rates of 1.71 % in 1820-70 and 2,14 % in

1870-1913 compared to 0.42 in 1500-1820. And it grew at the average speed of later times, though the experience of the 20th century was much more volatile, with only 1.16 % p. a. growth in 1913-50, but a staggering 4.65 % p. a. in 1950-73 and still a respectable 2.03 % p. a. in 1973-98.

- In terms of GDP per capita, the break with the past is equally impressive: a growth of 1.00 % p. a. in 1820-70 and 1.33 % p. a. in 1870-1913 against a meagre 0.15 % p. a. in 1500-1820. And the growth speeds approach those of later periods, though again with a massive volatility in the 20th century (0.83 % p. a. in 1913-50, 3.93 % p. a. in 1950-73 and 1.75 % p. a. in 1973-98).

Note that this stylized picture is relatively robust with respect to the inclusion of further European and American countries:

- By and large, the Iberian peninsula and eastern Europe had a similar intertemporal pattern of growth, though with a somewhat later start in the 19th century and somewhat idiosyncratic developments in most recent times.
- All in all, the economic growth of the United States and other “western offshoots” as Maddison calls them (notably Canada, Australia and New Zealand) runs parallel to that of western Europe. As these were typically immigration countries, however, the upsurge of economic performance from the early 19th century onwards is, if anything, even more pronounced than in Europe.

For the rest of the world, the picture is much more differentiated, with GDP-growth being mostly due to population growth rather than a sustained improvement of economic performance as can be seen from the substantial gap between the growth of GDP and GDP per capita. Still then, at least Latin America and Japan appear to start a genuine growth surge after 1870, i. e. somewhat later than Europe and the United States. However, the later records of these countries diverge widely, with Latin America loosing track and Japan embarking on a forceful expansion in the third quarter of the 20th century, which is also the time when the rest of Asia begins a catching-up that has much accelerated in the most recent past up to the present.

All in all, the data make abundantly clear that, roughly parallel to industrialization, economic performance changed dramatically. It is fully justified to speak of a fundamental change in terms of economic growth. We do not much exaggerate when calling it a change from statics

to dynamics. The positive drama of this change can best be illustrated by taking a side-look to measures of the absolute index levels of real GDP and GDP per capita before and during the time of industrialization. Figures 28 and 29 present Maddison's estimates of such levels, which were actually used by him as the basis for the annual growth statistics presented in Figures 26 and 27. Note that they are measured in terms of 1990 international Dollars, which involves an extremely tedious effort in estimating the purchasing power parities across countries and over time.

The results fully confirm the positive dramatics of the change: in 1820, real world GDP was 181 % higher than in 1500, and GDP per capita only 18 % (!) higher. Hence, the largest part of what was there in terms of (slow) growth has to be ascribed to population increase. For "core" western Europe (in the tables the so-called 12 Countries Total), the performance was better, but not very much so, with per capita GDP rising by a total of 60 % over 320 years from 1500 to 1820. In the following five decades from 1820 and 1870, the same group of countries reached a 64 %-increase of GDP per capita, and in the just 43 years from 1870 to 1913 another 77 %-increase. At that time, "core" western Europe's real GDP was 5.8 times its level of 1820, and GDP per capita about 2.9 times. In less than one century, much more was achieved in terms of growth than in the preceding three centuries, and maybe more than in any previous period of history. Clearly, industrialization opened the door to a new dimension of economic performance.

This is a lecture on economic Globalization and world economic development, not on economic growth per se. To be sure, historians agree that it was in the course of the 19th century that a first pattern of genuine Globalization took shape, parallel to unprecedented economic growth. As we shall see, this pattern cannot be understood without a thorough prior understanding of how the forces of industrialization changed the economics of the western world. This is why, in section 3.1, we start our analysis on the supply side of the economy and then move on to the cross-country flows of goods and mobility of production factors. In the first three sections (3.1.1-3), we draw a stylized picture of the different stages of industrialization and their consequences for technology, costs and the division of labour, with a particular view to long-distance transport and communications. In the last three sections (3.1.4-6), we then analyze how the trade of goods, the movements of capital and the migration of labour developed against the background of an industrialising world.

Note that, from now on, we mostly speak of "international" and not anymore of "long-distance" trade. Likewise, we speak of "international" movements of capital and labour. This is a tribute to the fact that, from the 19th century onwards, the adjective "international" –

though still far from analytically precise in economic terms – does at least begin to make sense and convey the right message. Nevertheless, it is still important to keep in mind that, even in a world of ever more well-defined nation states as it developed in the period under consideration, national boundaries are generally not economic ones. Hence, even in modern times, economic Globalization has retained an important intra-national dimension, at least when it comes to relatively large countries.

3.1.1 Industrialization

Historically and technologically, the process of industrialization is enormously complex. For our purposes of analyzing Globalization and world economic development, however, it is important to identify those few major changes that had far-reaching consequences for the performance and the structure of national economies. We therefore draw here an extremely stylized picture of the process.

That picture must start from the most fundamental fact that distinguishes “industry” from pre-industrial “crafts” and “manufacturing”: the size of the installed capital stock. We speak of a genuine industry if (and only if) it involves the substantial use of machinery, in contrast to mere tools. In an industrial factory, workers and employees do not just use small-sized tools as a kind of extension of their physical and mental powers. Rather, they work together in large numbers, and they do so with a complementary stock of machinery that massively raises their marginal and average labour productivity beyond the level that they would have if the machinery were not there. Economically, the machinery is of course nothing than prior work, skills and knowledge transformed – through the act of saving and investment – into a physical object, a “production detour” (a translation of the German term “Produktionsumweg” coined by the early 20th century Austrian economist Eugen Böhm-Bawerk). But it is precisely this detour that leads to the massive gains in labour productivity.

Of course, before industrialization, there had been many investments in various forms of capital stocks as we have seen in the course of the lecture so far. And, in ancient times and again since the Middle Ages, there was always a substantial capital stock available in the economy. Examples are livestock and ploughs in agriculture, ships, carts and draught animals in transportation, water- and windmills in energy production, and manifold tools in crafts and traditional manufacturing. What was generally lacking, however, were large-scale machines that were used complementarily to a large number of workers.

Note that this was true even for the famous “Arsenal”, the big medieval shipyard of Venice, which is today considered as the first plant of modern industrial size. Though indeed large, the Arsenal was pre-industrial in the sense that it did not make labour work together at a few big machines. Rather it was like a giant manufacturing plant in the true sense of the word, i. e. many craftsmen working with their tools, skills and hands (in Latin: “manu”) in one place at different parts of the product to be constructed (the “ship” or “boat”). Hence the Arsenal was an example of an efficient organisation of crafts rather than a genuine industry. Even as such it was, of course, altogether exceptional at that time, but it was not a forerunner of “industrialized” manufacturing.

Installing large-scale machinery can only be profitable if all complementary factors and all inputs are available in the place where production is planned. In the historical process of industrialization in the 19th century, the bottlenecks were primarily of two kinds, labour and energy:

- **Labour:** In a still rural society with very few and mostly small cities, labour was widely scattered over the countryside, working in agriculture or petty crafts and trade at low productivity and for low wages. Given the technical constraints of the emerging large-scale machinery, there was practically no alternative to workers moving to the capital stock (and not the other way round!). Due to its high capital intensity, the new industry could pay much higher wages than agriculture, and that led to a drain on rural labour towards the new industrial centres. In this sense, the trend towards urbanisation that could first be observed in the Middle Ages repeated itself, though on a much more massive scale. As we have described in section 2.2.2.1, this did not mean that, in the end, everybody was better off by migrating. In fact, due to massive crowding, the sanitary conditions in the industrial centres soon became dreadfully bad so that many migrants may have ended up at a lower standard of living than the one they had before. Still then, the “option value” of migrating was high enough to keep up the rural flight of labour for a very long time. Given the high birth rates in rural areas, the perspective of remaining stuck on the land was bad enough to make many move to the cities. No term describes this situation better than Karl Marx’ dictum of the “industrial reserve army”, which stood ready in the rural areas to fill whatever job slots emerged in urban industry.
- **Energy:** While, by itself, complementariness of capital and labour led to a powerful urbanisation, it was the need for energy that channelled this urbanisation in a direction that turned out to be quite different from the structure of the Middle Ages. At that

time, the nodes of commerce and trade tended to be the targets of migration. In the early 19th century, however, the running of the new industrial machinery required energy supplies that by far surpassed anything that could be delivered from traditional sources like water or wind mills or charcoal from timber. As coal was still very expensive to transport, it was natural that industrial growth concentrated in regions with readily accessible coal supplies either through seaports (like London in Britain or Hamburg in Germany) or, more frequently, through mining in the immediate neighbourhood. This trend was all the more pronounced the more energy-intensive an industry produced, with iron and steel manufacturing leading the pack. This is why the massive 19th century industrialization went hand in hand with a massive urbanisation with a strong bias towards coal-rich regions. In fact, within a relatively short period of time, the European map of urbanised regions began to change significantly: those areas with coal supplies like, e.g., the Ruhr valley in Germany, the Tyne-region in Britain, Lorraine in France and Wallonia in Belgium became densely populated agglomerations. As early as 1840, the new resource-biased pattern of urbanisation can clearly be recognized on the map of Europe (see Figure 30).

It is remarkable that, from early on, this new wave of urbanisation had a tendency to feed upon itself. Once vast numbers of rural families had moved into the new industrial centres, the sheer extent of labour supply in these regions led to the attraction of further industries. In later stages of industrialization, e. g., the manufacturing of textiles and clothing tended to cluster at those places where many women (and children!) stood ready to work in the factories. Quite naturally, this was the case wherever the male labour force had jobs as well, e. g., in coal mining and the iron and steel industry. This is why the German Ruhr and lower Rhine valleys became centres of both, “heavy” and “light” industries, with the light ones in a sense following the heavy ones.

It is hard to overestimate the economic, political and social significance of industrialization-cum-urbanisation in the 19th century. By its very nature, this process had a whole series of important consequences that have shaped the world until today. At this point, we pick out five of them, which are of particular importance for our further proceedings:

1. the trend towards trade,
2. the growth of labour unions,
3. the birth of unemployment,
4. the accentuation of the business cycle, and
5. the rise of the welfare state.

→ 1: The trend towards traded goods

Industrialization entailed a sustained and systematic bias towards trade. To see this, we have to develop a highly stylized picture of growth and structural change induced by industrialization. We start with the strictly macroeconomic framework of an aggregate production function that we already used in section 2.2.3 when describing different phases of the Middle Ages. In this framework, industrialization may be interpreted as a massive rise of the economy's total capital stock combined with a massive improvement of technical knowledge. Both together lead to a sharp increase of the marginal and average productivity of labour at given levels of employment, which then allows for a rise of the real wage combined with an increase of employment.

This is described in Figure 31, which shows a standard graph of a production function in labour/output-space. Figure 31 is constructed like Figure 5, and the underlying model has exactly the same properties as the one developed in section 2.2.3. The “exogenous shock” introduced by industrialization is a big upward shift of the graph as depicted in Figure 31. Of course, a long series of subsequent upward shifts would be a more realistic graphical representation, but for the sake of argument, one big shift will do. What happens is that, at a given level of employment L^* , the marginal productivity of labour (i. e. the slope of the curve) increases sharply, and there is scope for a movement along the new curve, which – depending on the location of the new equilibrium – means a higher real wage rate and/or higher employment. When the shift is big enough, one would expect scope for both: a higher real wage and a higher level of employment. In macroeconomic terms, this is in fact what we observe all over, and it marks the big difference between the expansion in the Middle Ages and that of the 19th century. In the period 1000-1300, there was quite an economic expansion with capital accumulation and technical progress, but it is not clear whether labour profited from it, as we have argued in sections 2.2.2 and 2.2.3. In the period 1820-1913 there was an even much more forceful expansion, and there is no question that labour did profit from it, at least in the longer run.

To be sure, the pure macroeconomics – though no doubt useful – masks some extremely important structural changes that deserve our attention because they have much to do with trade. To see this, think in terms of a simple three-sector world consisting of “agriculture”, “manufacturing” and “services”. History tells us that the “exogenous shock” of capital accumulation and technical progress took place mainly in manufacturing, much less so in agriculture and probably least in services, which continued to work (like crafts) with relatively simple tools. Think of the traditional hairdresser or barber who had hardly the opportunity to substi-

tute a big machine for his scissors or razor-blades. What happened historically is that a massive “flight” from agriculture into industry took place, but not from services into industry.

How can this be explained and what significance does it have? The answer is graphically contained in Figure 32, which depicts demand and supply curves (denoted D and S) in quantity/price-space for a typical industrializing country in the 19th century with three sectors: agriculture, manufacturing and services. As usual in this lecture, the figure is to be regarded as purely illustrative to carve out the main line of reasoning. Note that the supply curves of the three sectors have a standard positive slope: the higher the output price, the more they supply (other things being equal). For the sake of argument, the demand curves for the three sectors are drawn very differently: a horizontal line for agriculture and manufacturing, which are assumed to produce tradable goods in perfect competition with foreign suppliers so that the home market price is determined in the world market (the standard “small country assumption”); and a vertical line for services, which are assumed to be completely non-tradable and face a price-inelastic home demand (e. g., the haircut). Also for the sake of argument, we assume perfect mobility of labour between the sectors: labour moves to the sector that pays the highest real wage.

Let us now assume that, in manufacturing alone, new machinery is installed, which incorporates a massively improved technology so that the supply curve is substantially shifted to the right. This means, in a general equilibrium context (which is of course not fully worked out here) that, due to resource constraints, the supply curves in the two other sectors move to the left. This leads to lower production in agriculture, but no change of production in services. Why? Because agricultural producers are unable to pass higher labour costs onto consumers while service producers can do exactly this as they offer an output that cannot be bought abroad or easily dispensed with. Hence, as a consequence of industrial productivity growth in a framework of intersectoral labour mobility, agriculture shrinks and services become more expensive.

Note the most interesting result of the original exogenous shock. The value productivity of labour and the real wage increase everywhere, but for different reasons:

- in manufacturing because physical labour productivity rises (the original “shock”) and allows to attract labour at higher real wages,
- in agriculture because labour leaves for industry and thus raises the physical labour productivity of the remaining workers, and
- in services because – at unchanged physical labour productivity – service consumers are ready to pay more for the output, thus allowing prices to rise.

The last effect is well-known in the theory of international trade as the Balassa-Samuelson-effect, which explains why prices of non-tradable goods (usually services) are positively linked to physical productivity in the traded-goods sector (usually assumed to be manufacturing). Note that this effect has much power in explaining why services in non-industrialized (and usually poor) countries are much cheaper than in industrialized (and usually rich) countries so that international comparisons of living standards on basis of official exchange rates tend to overstate the real income gap between rich and poor. To arrive at reasonable results, one has to use purchasing power parities, which take account of this effect.

For our purposes here, however, something else is of greater interest, and that has to do with trade: a structural change in the course of industrialization as we have described it in this simple model has an inherent quantitative bias towards extending the volume of international trade. Why? Because the growth induced by industrialization raises the value, but not the volume of non-traded goods. On the other hand, the movement of resources from agriculture to manufacturing tends to increase the volume of trade: the country that experiences the positive “shock” described in Figure 32 will tend to produce and export more manufactures and produce less and import more agricultural products, which is roughly what happened in the 19th century in those countries that did in fact industrialize fast.

Note that the “trade bias” of industrialization comes about independent of what exactly are the determinants of trade flows. Standard trade theory delivers a variety of reasons why a country might increase its production and exports of industrial products and increase its imports of agricultural produce. E. g. it does not matter whether the comparative advantages that determine trade flows are due to relative differences in intersectoral labour productivity (Ricardian theory) or in factor endowments (Heckscher-Ohlin theory).

Note also that the very simplicity of the model, if anything, understates the scope for profitable trade in the course of industrialization. A more realistic model may further split up the industrial sector itself and allow for economies of scale and product differentiation in a setting of monopolistic competition in world markets. If this is done, the realistic conclusion will be that industrialization does not only enlarge the scope for inter-sectoral trade (e.g., manufacturing goods for agricultural produce), but also for intra-industry trade (say, steel for textiles or machines for other machines).

Sure enough, reality is anyway much more complex than the model and the above reflections suggest. In particular, shrinking sectors typically react to higher labour costs by mod-

ernising their equipment, which helps to raise labour productivity. In fact, the 19th- and 20th-century history of agriculture in the industrializing regions of Europe, America and elsewhere is a never ending sequence of mechanisation: in the face of rising labour costs, agricultural producers invested in new machinery, thus gradually turning agriculture into a kind of agro-industry as can impressively be observed in the most modern types of farming that emerged at that time, e. g. the highly profitable sugar-beet plantations on the extremely fertile loess soils in the West of Magdeburg (the “Magdeburger Börde”) and Cologne (the “Kölner Bucht”). Of course, this did neither prevent the shrinkage of agriculture in terms of employment nor did it substantially slow trade integration both in manufacturing and (modernising) agriculture. However, it led to a spreading of productivity advance from industry to agriculture and thus in the end to mechanisation penetrating the whole economy (except personal services).

→ 2: The Growth of labour unions

Industrialization-cum-urbanisation had an obvious effect on the distribution of labour in space: it led to a geographical concentration – in large plants at large machines where the workers worked, and in large cities where workers lived with their families. This paved the way for the creation of labour unions that stood ready to pursue workers’ interests in higher wages and/or better working conditions through collective action, from wage bargaining to strikes. In a way, labour unions became for industrial workers what guilds had been (and continued to be) for artisans and craftsmen: cartel-like arrangements to further their interest.

Of course, geographical concentration may be a necessary condition for successful unionism, but it is not a sufficient one. Sure enough, it tends to allow intensive communication and coordination among a large number of people: meeting each other at work and living close to each other in a borough of a big city certainly facilitates the formation of joint aims and goals as well as the organisation of the means to achieve them. Economically, however, closeness is not enough for successful collective action. There must also be a strong market position of the unionised workforce, which may either be due to specific indispensable labour skills and/or to a technical complementariness of labour to large-scale machines that makes it possible to stop vast lines of industrial production by a well-coordinated strike of many workers, thus potentially inflicting a massive cost to the employer.

Given these economic pre-conditions for successful collective action, it is not surprising that unionism started its rise in those occupational groups of labour like printers whose high skills and indispensable place in the production process came closest to giving them a strong monopolistic power vis-à-vis employers. However, the further industrialization proceeded, the

stronger the element of technological complementariness became: ever larger machines were used and mechanisation made for an ever stronger interdependence of stages in production within firms so that an ever broader class of workers had a strong incentive to form cartels and pursue collective action. E. g., in Germany, the union of metal workers in the end became the largest – and disputably most powerful – union in the world, as it had members in all those highly capital-intensive branches of metallurgy, metal manufacturing, machine building and mechanical engineering that were to form the backbone of the German economy in the later stages of industrialization and beyond.

To be sure, the development of unionisation took time and it met resistance, not least by governments, which tended to prohibit or at least hinder union formation and collective action as it was typically linked to social democratic or other left-wing political groupings and parties that were on a parallel rise since the mid 19th century. Also, employers began to found their own associations so as to coordinate and counter industrial action by unions, e. g., through lockouts that raised the cost of strikes beyond what unions could finance on the basis of the contributions of their members. Thus, gradually, all industrialised countries developed a kind of network of “corporatist” institutions of industrial society, which were not confined to the labour side and not just exclusively concerned with issues of the labour market.

Naturally, these developments were highly differentiated between countries: strong in Britain, Germany and later the Scandinavian countries and somewhat weaker in the Romanic or Mediterranean part of Europe and the United States with its strong free-market tradition. In all countries, however, the trend was the same, and for basically the same reason. Not before the degree of industrialization passed its peak some time in the second half of the 20th century did the power of unionism begin to decline. Even today, however, unions – and also employers’ associations – are major collective actors in virtually all countries of the industrialized part of the world.

→ 3. The birth of unemployment

Until industrialization took off in the 19th century, there was no “unemployment”. In fact, the word itself did not exist, and neither did an idea of what it could possibly mean. To be sure, a pre-industrial rural society knows many sorts of social calamities: states of misery and poverty, of hunger and disease, of starvation and devastation, but it does not really know a state of unemployment in the sense of many people who are willing to work and offer their labour do not find a job.

Why is this so? The answer is straightforward: In a basically rural environment of largely self-sufficient agriculture, there may be fluctuations in general economic conditions on the demand and the supply side, but these normally induce no more than fluctuations in the degree of economic activity – and not unemployment in the modern sense. The simplest example is the seasonal character of agricultural work: in a technical sense of capacity utilisation of labour, farmers and their families tended to be vastly underemployed during winter time, but they did not perceive it as a problem because the intensity of work could simply be reduced. The spare time was then used for whatever supplementary activity outside the main production line of agriculture such as work in the household or repairs.

A similar adjustment took place in times of non-seasonal fluctuations of demand and supply: bumper harvests simply required more work and poor harvests less. Only when harvests turned out extremely bad over a longer period of time, migration set in, and then at times on a massive scale. In the Irish famine in the 1840s, the threat of starvation led to mass migration – in the Irish case to British and American cities. Even then, however, these mass movements of people were the consequence of rural poverty rather than of anything that could reasonably be called “unemployment”.

In fact, unemployment only emerges when there is neither the option of varying the intensity of work nor, in the extreme emergency case, the option of migrating. And this was typically so in those fast-growing industrial cities that were themselves the target of migration in the first place. Once people became settled in a big 19th century industrialised city, they had no real option anymore to leave for somewhere else in case of unemployment. And once an industrial worker and city dweller was laid off from a big manufacturing plant for whatever cyclical or structural reason – say, a slack of demand in the industry or a bankruptcy of the firm – he had to find a job somewhere else in the city. One or two generations after migrating from the countryside to the city, there were no links anymore to the rural roots, and a return out of the question – not least because, in normal times, the city offered more productive work and better wages. Hence, paradoxically, by functioning like big job machines, the industrial cities themselves became the cradle of the modern phenomenon of (involuntary) unemployment.

By and large, the industrial growth dynamics of most of the 19th century ensured that urban mass unemployment remained a strictly temporary phenomenon – confined to those periodic, but relatively short cyclical crises that were typical for the newly emerging industrial capitalism (see below). This is why other social problems than unemployment played a much more prominent role in the political discussions that emerged in the newly industrialised part

of the world in the second half of the 19th century. Still then, however, cyclical unemployment was clearly recognised as a new type of risk and danger, and in some countries like Germany, it motivated unions to offer rudimentary schemes of contribution-financed unemployment insurance.

→ 4. The accentuation of the business cycle

More generally, the emergence of cyclical unemployment leads straight to another consequence of industrialization: the new power of the business cycle. To be sure, cyclical fluctuations of economic activity had always existed. However, the greater the relative weight of industry in output and employment in an economy, the more pronounced they became.

Why was this so? The answer lies in the characteristics of most industrial production – as compared to agricultural output and also personal services. Until the very recent past and certainly in the 19th century, most of the farmers' output served as food for humans or animals. Naturally, there are regular fluctuations on the supply side, notably the four seasons and the stochastic incidence of good and bad harvests. But normally, they do not lead to very pronounced business cycles because the “game” of seeding and harvesting is regularly repeated so that disturbances remain relatively short-term phenomena. On the demand side, there are changes of tastes and preferences for food over time, but they tend to be very slow and gradual, and they do not really have a strong cyclical dimension. After all, eating and drinking is a regular and relatively constant activity, both for humans as for animals. The same holds for the demand for standard personal services (e. g. the hair-cut).

Things radically change once an economy starts producing a broad array of sophisticated investment goods and of non-perishable consumption goods. Here, the determinants of demand tend to be highly complex and interdependent. Economically, the purchase of an investment good or a long-living consumption good has the character of a “stock adjustment” – in sharp contrast to the “flow” of consumption or production inputs. Typically, stock adjustments are based on a sophisticated calculus of intertemporal optimisation: a commercial investor who invests in new machinery or a private family who “invests” in long-term household equipment do so only if they find it to be rational and profitable, i. e. if they expect a (properly discounted) stream of returns or services from the use of the respective capital stock that is higher than the cost of the purchase.

It is quite obvious that this very profitability in the broadest sense is much more dependent on general economic conditions, i. e. on the decisions of other economic agents that make

similar dispositions. This is notably so in business: e. g., an investment in a machine that is used in the production of tools which are applied in other branches of manufacturing will depend on the demand conditions in these other branches, and will fluctuate accordingly. Clearly then, capital goods industries in the broadest sense (e. g. including the production of private household equipment) are typically subject to a high degree of correlation of their market conditions, which will add an element of cyclical autocorrelation to growth.

Less technically speaking: the price of dynamic industrial growth in the 19th century and beyond was (and still is) a wider fluctuation of business activity. This is why an industrialized world is also typically a world of powerful business cycles, and the economic history of the 19th and 20th century is in fact full of them. We shall repeatedly come back to up- and down-swings of business as well as to full-scale booms and recessions in the course of our further proceedings.

→ 5. The rise of the welfare state

Industrialization-cum-urbanisation changed the world not only economically, but also socially. A static, poor, but relatively safe rural world was replaced by a dynamic, more prosperous, but rather dangerous one. And the new risks involved could not be insured in the traditional form by large families and possibly the village community that provided shelter and protection in the case of individual hardships. To be sure, most large rural families were desperately poor in pre-industrial times so that the family network of insurance was not really a very comfortable one. In fact, one should beware of any romantic overtones when looking at the “good old times” of traditional family and village networks, not to speak of the restrictions on freedom and personal responsibility that they entailed. Be that as it may, however, the traditional networks did provide some protection, which gradually faded away in the course of industrialization and urbanisation.

Step by step, this fundamental social change was recognised – and so were the abominable medical and social conditions in the big urban centres of industrialization. From the middle of the 19th onwards, a lively political and academic debate took place in the most advanced countries about what was later in Germany called the “social question”. In one country after another, this led to new public regulation of industrial activity (e. g. restrictions on child labour) and to legislation on social security matters. This trend gradually led into something that may be called a rise of the welfare state. While further extensions of social security networks in European countries and, to a lesser extent, in America took place in later times, in

particular the 1960s and 1970s, major first steps were made in the later phases of 19th century industrialization. Notably Germany became a “pioneer” of the welfare state when, in the late 1880s, Bismarck’s government established the first public social security system of the world. Note that much of this legislation – and also Bismarck’s – was not really motivated by a genuine concern for the miserable and poor people. In large part, it was a merely political instrument to pacify a labour and union movement that was growing ever stronger.

So much for the major long-standing consequences of industrialization in five selected fields. It is important to realise that the process of industrialization and its consequences have never been fundamentally reversed in those parts of the World where it gained ground in the 19th and early 20th century. It is true that, in recent decades, there have been trends towards what is often imprecisely called a de-industrialization – meaning that the complementary machinery was split into smaller units and modern computers and information networks allowed new forms of decentralised work. However, these new trends did never mean that installed machinery and capital stocks lost their importance for labour productivity in industry and to some extent also in services. And they did not remotely reach the power of the forces of industrialization, and did certainly not mean anything like a return to the degree of decentralized economic activity that had prevailed in pre-industrial times. We shall return to these important matters of structural change in the section 3.3 of this manuscript and part II of the lecture.

3.1.1.1 Britain’s Lead 1820-50

Britain was the first European country to industrialize. Whether this happened in the course of an extremely fast “industrial revolution” or a somewhat slower “industrial evolution” has been the subject of a long-standing debate among economic historians of the Anglo-American academic world. For us, this is a not a very interesting question because it concerns details of the time pattern of change and not the facts of the change itself.

For us, it suffices to state: industrialization took place first in Britain, and it began in the second half of the 18th century, accelerated markedly after the Napoleonic wars and reached something of a first peak in the middle of the 19th century, with the Crystal Palace world exhibition in London in 1851 being a wonderful historical mark for defining a kind of maximum distance between the leader Britain and the followers on the continent and in America. That distance could be kept constant for roughly two decades, but then began to decline sometime after 1870 when the catching-up of other countries gained momentum. On the eve of World War I, the lead had melted down, though not completely disappeared: in terms of per-

capita income, only the United States had actually overtaken Britain, but all major European countries had come much closer to it (see FIGURE 29).

With the benefit of hindsight, it is no surprise that Britain was the first country to industrialize: a society open to new ideas, an economy open to world markets, a relatively strong commercial and mercantile tradition, a growing community of able engineers and entrepreneurs, a remarkably rational public debate on economic policies, and – not least – the vast energy supplies favoured the British isles over the rest of Europe, even over Holland, which shared many aspects of modernity with Britain, but had no coal in the backyard. Note, however, that without the benefit of hindsight, it would have been much more difficult to forecast the British success, say, in the 18th century because the venture into an industrialized world was something altogether new, with no historical precedent whatsoever.

As early as 1800, a clear regional pattern of industrial activity in Britain had become visible (see FIGURE 33): the cotton industry in Lancashire, where raw cotton was imported from America via the port of Liverpool and then manufactured in and around Manchester; the wool industry around Leeds; iron and steel close to coal mining in the regions of Newcastle, Birmingham, Coventry and Cardiff; cutlery in and around Sheffield; pottery in Staffordshire, etc. In fact, these are precisely the industries that were the most dynamic in the first half of the 19th century, which led to the geographical structure of urbanisation that we still see today in Britain.

Note that, right from the start, these industries could rely on a remarkable transport infrastructure, notably a dense network of canals and turnpikes that had been built or were up to be completed in the late 18th and early 19th century (see FIGURES 34 and 35). This network was able to carry large quantities of bulky, low-value goods to an extent that was yet unthinkable in most other countries. It was created by a mixture of private initiative and public chartering, which allowed to raise large amounts of private capital, but at the same time ensured the necessary minimum of public guidance and surveillance so as to add up to a well-structured system of infrastructure. Hence, even before the advent of the railway, British industry had access to a remarkably efficient transport network.

Parallel to industrial expansion, population grew fast. Between 1750 and 1800, Britain's population rose by 45 %, and from 1800 to 1850, it almost doubled, from 10.7 to 20.9 Mio. people. It is precisely in this first half of the 19th century that Britain – for the last time ever since – topped all European countries in terms of population growth. It is at that time that Britain

became one of the most densely populated areas in Europe, which it had never been before, but remained ever since.

3.1.1.2 Others' Catching Up 1850-1914

A casual glance back at the FIGURES 26 and 27 shows that, from 1870 on, most European countries and the United States had faster economic growth than Britain, and some quite substantially so. Notably the economies of northern and central Europe expanded fast, and they did so by making a big leap in terms of industrialization.

Much effort of economic historians has been put in answering the question why Britain's lead narrowed and why other countries caught up substantially at that time. Why was the first and the second part of the period 1820-1914 so much different in relative economic performances? For our purposes, the question itself may be too ambitious. After all, why should one expect that the British economy was bound to gallop away from the rest of the pack and then conserve the lead forever? To be sure, technical knowledge spread over national borders, and this was not only so for reasons of industrial espionage, which was very common at that time. In many respects, there were simply too many structural similarities between Britain, northern, western and central Europe and also Britain's overseas offshoots that one should have expected anything else than some sort of convergence to a new and much higher standard of living, once industrialization had run its further course.

To be sure, the convergence that happened in Europe between Britain and the rest of the continent up to the first World War was a strictly partial one. FIGURE 36 shows a measure of real GNP growth for the period 1830 to 1913 for twelve European countries (based on older statistics that have been used by the American historian Norman Pounds, not by Angus Maddison, though the difference is not relevant for the purpose at hand). FIGURE 36 shows that the accelerated "catching-up" growth after 1870 was strictly confined to central and northern European countries. It did fully exclude the Mediterranean, with the possible exception of Italy, which belatedly started on a growth spurt that was largely confined to the northern parts of the country. All in all, it is clear that, on the eve of World War I, a pronounced North/South-development gap had emerged in Europe, which was to survive almost all over the 20th century right up to the most recent present.

Of all large countries, the most striking case of catching-up was that of Germany (the fat line in FIGURE 36). In terms of per-capita income, it had overtaken France by 1913 (see also and more precisely FIGURE 29), though not yet Britain, Belgium and the Netherlands. How-

ever, due its large (and fast growing) population size, it had the largest national economy on the eve of World War I (see FIGURE 29): about 6 % larger than that of the United Kingdom and 64 % larger than that of France, the only of the larger European countries that did not experience a strong increase of its population in the 19th century. Clearly, the pace of German industrialization was massively accelerating in the later 19th century: by 1913, its employment share of industry reached 36,3 %, higher than in any continental country except Belgium and Switzerland.

Statistics on industrial inputs are even better in giving a taste of the fast and thorough transformation of Germany into an industrial economy and society (FIGURE 37). From 1840 to 1910, there was a rise in consumption per capita of raw cotton from 0.9 to 6.8 kg, of pig iron from 5 to 220 kg and of coal from 110 to 3190 kg. Hence, in terms of pig iron input, Germany eventually overtook Britain, in terms of coal input, it came close to it, and in terms of cotton input, it substantially narrowed the gap to Britain, which was by far the greatest cloth and textile producer in the world. Note that, in terms of total output of pig iron and steel, Germany even surpassed Britain by a wide margin (see FIGURE 38).

With the catching-up of central and northern Europe – and notably Germany –, the European continent began to exhibit the pattern of urbanisation that is familiar to us until today. FIGURE 39 shows an urban map of Europe in 1910, with bigger cities drawn as fat dots of varying sizes and the average national share of urban population being marked by different shadings. The message is unambiguous: all those urban and industrial centres that we still know today did exist on the eve of World War II, with the strongest concentration to be found in western and southern Germany, Belgium, the Netherlands and northern France (and, of course, beyond the Channel in Britain, which is excluded in the picture). In the rest of Europe, there are also big cities, but they tend to be scattered much looser over space, thus indicating a much lower degree of urbanisation and industrialization. Note that the shading of population density in FIGURE 39 broadly supports this view.

In the early 1990s, the American trade economist Paul Krugman published a slim, but important book with the title “The Geography of Trade”. In this book, he argued convincingly that industrial and urban patterns – once established – tend to have a very strong tendency to remain as they are. They are inert or path dependent: once people have moved to the places where industry develops for reasons that might be geological, technological and/or geographical (e.g., coal fields that lower energy costs or rivers that lower transport costs), the spatial concentration of economic activity tends to continue even if the specific locational cost advantage loses its importance. From a casual glance at the map of Europe in FIGURE 39,

we can draw the conclusion that there is a lot to Krugman's argument; and we would discover the same type of phenomenon if we looked at a map of North America or maybe Australia. Even today, when coal and rivers have lost much of their importance, the geographical pattern of urbanisation roughly remains as it developed at the times of industrialization. Apparently, in later periods, the new industries and services that grew came to the people, and not the other way round. In this sense, something important has changed.

What is it economically that has changed? Without going into any details, the key to an explanation probably lies in the working of external economies of scale that tend to conserve spatial agglomerations of people and economic activity over time. E. g., once an industry – say, iron and steel – had its stronghold in the Ruhr valley, specialized suppliers of inputs clustered there as well, which facilitated the growth of other branches of metal manufacturing that needed largely the same inputs, though much less coal. Similarly, a highly skilled and specialized labour force of metal workers and craftsmen made up what is often called a "thick" labour market – meaning the availability of an ample supply of specialized skills at short distance, which is also an advantage for related, but less coal-dependent new branches of manufacturing. Finally, the centres of industrial research and innovation in metal manufacturing also developed in established iron and steel regions, thus facilitating off-springs in new (and less energy-dependent) fields to develop. Thus, almost ironically, the Ruhr valley today has become a centre of modern environmental technology – not because it was particularly "clean" in the first place, but because there are strong linkages between some traditional technologies and new ones.

To be sure, these path-dependent industrial developments and their geographic incidence long after industrialization are among the most important characteristics of economic growth and structural change. Politically, however, they are often not conceived as a positive story, but rather as a crisis. After all, even if the "old" industrial region succeeds in restructuring, the public perception is one of a painful adjustment from a solid "good old" industrial base to uncertain future challenges. In terms of its macroeconomic record of growth, the adjustment will have no similarity to the dynamic expansion at the heyday of industrialization. And if it is accompanied by a significant rise in unemployment, which is often the case, the public disillusion will be all the more pronounced. Both economically and politically, we shall return to these issues later in the lecture.

3.1.2 R&D-based Industrialization

In at least one major respect, the followers of Britain – and notably Germany – added a new element of modernism to the process of industrialization. From about the 1870s on, more and more technical universities (in contemporaneous German terminology: “Technische Hochschulen”) were founded with the explicit aim at providing a high academic level of technical education. Gradually, these universities grew into the role of major providers of highly-qualified technical experts for industry. The theoretical curricula were more and more adopted to the needs of industry, with a strong emphasis on experimentation in laboratories and other methods of modern empirical research.

In Germany, the technical universities expanded rapidly in the last quarter of the 19th century and became an important source of “human capital” for those fast-growing branches of industry that had a strong need for a more scientific approach to innovation. In this respect, the paradigmatic example became the German chemical industry, a highly successful branch of manufacturing with staggering growth rates in terms of output and employment in the decades before World War I. FIGURE 40 gives a graphical impression of the tremendous expansion of the production of chemicals in Germany from 1850 to 1913 – in comparison to textile production. The graph makes clear that, within the German economy, the growth of the “modern” chemical industry by far outpaced the growth of more traditional textile production, which by itself expanded quite rapidly. Many today big chemical firms were founded in the Rhine/Main-valley, e. g. “Badische Anilin und Soda-fabrik” BASF at Ludwigshafen, Bayer near Cologne and Höchst at Frankfurt. And these proved to be highly world-market oriented firms with an increasingly scientific approach to research and development (R&D) of new products and processes.

Note that, in this respect, there is good reason to believe that Germany and, for that matter, other countries of continental Europe as well as the United States somehow “leapfrogged” British-style industrialization, which was based on a more traditional trial-and-error-approach to innovative new ideas. This may have given the followers an additional element of growth dynamics in the late 19th century, though this must remain to some extent a matter of speculation. In any case, the rise of technical universities in Germany spurred a political debate in Britain on copying the German model, which eventually led to the foundation of variants of technical universities in Britain as well. This very fact alone may indicate that, by that time, Britain had lost its lead: in terms of R&D-based industrialization, it was a follower and not a pioneer. Though it remained a wealthy country with a high standard of living, the powerhouses of innovative dynamics were now to be found in other places.

3.1.3 Steamships, Railways and Telecommunication

The 19th century was a time of most dramatic changes in transport and communication. In the first half of the century, steamships began to replace sailing ships on rivers and oceans, and locomotives, wagons and trains took over from horses with carts and carriages on land routes. In the second half of the century, ever more telegraph cables crisscrossed Europe and America. In 1866, the first transatlantic cable connected Europe and America, thus reducing the time of information flows between the continents from days and weeks to seconds and minutes. Both in transport and communication, there was a continuous flow of technical improvements that led to an equally continuous shift towards ever better machines, equipment and hardware of trains and ships. In fact, the technical museums of today give ample and fascinating evidence on this powerful process.

The new technologies in transport and communication led to a sharp rise of infrastructure investment. Investments took place in ports and canals, in railway tracks and stations as well as in telecommunication cables. The first wave came with the construction of artificial waterways and the improvement of shipping facilities (“canalization”) on natural rivers. FIGURE 41 shows, as a major continental example, the development of navigable waterways in the north German plains. Major canals were built and projected to link the river systems of the Rhine, Weser, Elbe, Odra and Vistula. And this, combined with technical improvements in shipping, allowed for a continuous rise of tonnage transported on rivers and canals. Between 1840 and 1900, the transport volume on these five rivers rose roughly by a factor of five, i.e. at an average compound growth rate of 2.7 %, with a marked acceleration in the last two decades (1880-1900), when the annual growth rate rose to 4.7 %.

Railway construction was even more impressive. Without much exaggeration, the second half of the 19th century may be called the age of the railway, as Figure 42 shows in a stylized fashion: by the end of the century, the rail had become by far the most important medium for long-distance transport – outperforming canals, rivers and also oceans, and for a while also the roads whose greatest time was to come in the 20th century with the advent of mass motorization. FIGURES 43, 44 and 45 show the stepwise growth of the European rail network: hardly existing in 1840, it had made considerable progress by 1850, notably in western and central Europe, and covered virtually the whole continent by 1880, though still with visible differences in density. The enormously fast growth of railway freight between 1850 and 1913

in virtually all countries (FIGURE 46) is proof of the all pervasive importance of this development for transport volumes.

As usual in economic history, there are relatively good statistics on issues that can be physically measured (length and location of railway tracks, tonnage of shipping and freight etc.), but not many pieces of hard statistical evidence on the economic effects in terms of cost reductions that were induced by all these innovations and infrastructure investments. What exists, however, does point to a most powerful effect. As to the cost of shipping goods across the Atlantic, economic historians have calculated freight rate indices (see FIGURE 47). They show that, between 1840 and 1913, transport costs in real terms (deflated by an economy-wide price index) declined by a total of 70 %, i. e. on average about 1.6 % p.a. per year, an enormous reduction indeed. Unfortunately, no such evidence for the changing cost of railway transport is available, but the sheer density and capacity of the emerging railway network must have led to dramatic improvements indeed.

For communication technology, no hard facts on costs are available either, but an idea of the sheer magnitude of change can be gained when looking at the development of price differentials in major financial markets. E. g., in early 1866, it still took about ten days to travel across the Atlantic, which was of course much less than at the times of the sailing ships, but it still meant that a message on financial markets that originated in the City of London took no less than ten days to reach New York's Wall Street (and vice versa). On July 27 of that year, the first reliable transatlantic telegraph cable started working, and the mean absolute price differential between London and New York plunged by a fat margin of 69 % over a few months – and then stayed constant for years. Within the United States, price differentials in financial markets generally declined by 25-40 % when telegraph cables were put to work. How much this meant in terms of cost saving for the economy as a whole must remain an open question. Clearly though, the actual gain in the use of information was enormous.

So much for the supply side effects of what may be called a genuine revolution in transport and communication technology. What is equally important, however, are the demand side effects. In fact, railway construction alone provided a massive demand stimulus notably for investment goods industries. For the first time in history, a completely new infrastructure stretching over long distances and whole continents was set up – and required tracks, locomotives and wagons that all needed iron, steel and metals as major inputs. Thus the enormous productivity potential of cheaper transport was transformed into a massive demand for investment goods that could be produced with an ever improving technology. Never before – and so far never again – has the need for a new infrastructure led to such a huge demand

push in a specific manufacturing industry that was technologically advanced by all standards of the time. It is therefore not at all surprising that steel and railroads were most fascinating objects of the imagination of 19th century-society, ranging from financial capital markets with their big railway manias (and occasional crashes) to the engineering hypes like the building of the Eiffel Tower for the Paris world exhibition of 1889. They shaped a whole age, and they left physical traces that have survived more or less intact until today.

3.1.4 International Trade

As to trade policy, the almost 100 years from the Napoleonic Wars to the eve of World War I can be subdivided into two clearly defined sub-periods: the path towards free trade, beginning in Britain in the early 1820s, then spreading to other countries in the 1860s and reaching its climax and turning-point in the late 1870s. Starting with Germany in 1878, a number of continental countries then turned to a more protectionist policy stance, which was to remain the predominant political practice until World War I.

Against the background of industrialization and massive cost reductions in transport and communications, this two-part record of trade policy makes the period 1820-1914 one of the most interesting to study major aspects of trade liberalisation and protectionism. What was the policy debate, and when, and why? What were the effects of liberalisation on the volume and structure of trade? How did free trade change the distribution of income? Who gained, who lost? Why did the path towards free trade eventually end in a backlash of protectionism? How serious was this backlash? Did it matter in a world of rapidly falling transport costs? Etc., etc.

In addition, the period delivers a wonderful playground for testing trade theories – at least to the extent that the uncontrolled experiment of history does ever allow to test a theory. This is particularly true for that most important piece of neoclassical theory that explains trade structures in terms of resource endowments and places them into a general equilibrium framework: the so-called Heckscher-Ohlin model. Very roughly speaking, this theory states that, under conditions of free trade, a country specializes in the production and export of those goods that are produced by intensively using that factor of production with which the country is abundantly endowed and which is therefore relatively cheap. And this pattern of specialisation then leads towards an equalisation of factor prices because it involves an international equalisation of factor scarcity: in all countries, abundant factors become scarcer, and scarce factors become more abundant, thus leading to potentially serious conflicts of interest due to a trade-induced redistribution of income.

In fact, 19th century trade integration is something of a canonical case for this theory because it involved a growth of trade in agricultural goods, produced land-intensively, and manufacturing goods, produced labour- and capital-intensively. And it should not come as a surprise that the pioneers of this important theory of trade, the early 20th century Swedish economists Eli Heckscher and Bertil Ohlin, had exactly the experience of 19th century trade integration in mind when formulating their theories, which were then later refined by a whole generation of modern economists, with the Nobel Laureate Paul A. Samuelson being the most prominent one.

More recently, it was the experience of the 19th century, which became the subject of an important historical study of Globalization, which was done with the relatively sharp tools of general equilibrium analysis and econometrics. In their book “Globalization and History – The Evolution of a Nineteenth-Century Atlantic Economy”, published in 1999, Kevin O’Rourke and Jeffrey Williamson, two renowned economic historians, tried to tackle quite a few of the questions asked above – and obtained a number of most remarkable empirical answers. The sections 3.1.4-6 owe much to their work.

3.1.4.1 The Path to Free Trade 1820-1878

As most wars of history, the Napoleonic wars that ended in 1815 were bad for international trade. In its persistent fight against Napoleon, Britain had subjected the continent to a sea blockade by the British Royal Navy, which effectively lasted from 1807 to 1813. On both sides of the Channel, this blockade – together with French counter action and the usual disruptions of war – led to a trend towards autarky: import-competing sectors profited and export sectors suffered, leading to a turning-back of an already established international division of labour. In France (and other parts of Europe), it was the inland manufacturing sectors that profited from war-time protection, which may have laid the ground for a long-standing protectionist tradition in manufacturing branches on the continent that came to the fore again in the last quarter of the 19th century. In Britain, the main profiteer was agriculture and the landlords. They obtained high prices for their grain as the traditional main supplier, which was still Prussia, took part in the wars, though towards the end as an ally of the British.

With the wars ending in 1815, grain prices were bound to collapse – and with them the comfortable economic position of the British landlords who were politically influential in the conservative Tory party. As a result of their lobbying for protection, the famous “corn laws” were passed in 1815 (the term “corn” in English meaning “grain” or “wheat”, not “maize” as in

American usage!). The corn laws stipulated that wheat could be physically imported, but not sold in the domestic market if the price of corn was lower than a certain threshold level, which was 80 shillings per so-called quarter. This was high indeed since the price had been about 74 shillings in 1814 (still at war time!) and 52 shillings in early 1816. As a matter of fact, it led to a complete exclusion of foreign grains from the domestic market for seven years.

Gradually, a political demand for less protectionism began to gain ground in the public, with the economist David Ricardo being in the forefront of discussion until his early death in 1823. In his famous book “Principles of Political Economy and Taxation”, first published in 1817, he made a forceful case for free trade on the basis of the so-called comparative advantage of nations, which he demonstrated in a model that is still today a standard reference in textbooks of trade theory. Note that his model was brilliantly articulated in theoretical terms, but not a very realistic description of reality because it only used one factor of production, thus elegantly by-passing the typical distributional effects that are involved in free trade when more than one production factor exists. It is speculation whether, at this point, he was naïve or particularly clever as a proponent of free trade who wanted to avoid providing his opponents with a strong distributive argument against trade. Be that as it may, his unambiguous argument for trade prominently entered the public debate and became an intellectual fore-runner of later liberalisation. In a sense, Ricardo can be taken as the first advocate of free-trade Globalization, just as, one generation before, Adam Smith had become the first advocate of a market economy.

Ricardo's theoretical arguments found political support among urban manufacturers who naturally supported free trade. Politically, their interests were considered by the Whig party, the forerunners of the later liberals. In fact, step by step, the corn laws were liberalised. In 1828, a sliding-scale tariff replaced the conditional sales prohibition – thus at least allowing grain imports, though with a specific tariff that was set the higher, the lower the domestic price happened to be. Further tariff reductions followed until eventually, in 1846, the corn laws were repealed. Sure enough, the final repeal was more the final result of a long evolution rather than a revolutionary new policy. In fact, it has been estimated that the effective tariff protection – calculated as the ad valorem equivalent of the specific tariffs levied on grain – declined from a hefty 70 % over the period 1815-27 to 50 % in 1828-41 down to 7 % in 1842-5 and zero from 1846 onwards.

Note also that the fight against protectionism was not an isolated process in British society. To be sure, there was a very effective political lobbying for free trade in the later 1830s. A

breakthrough in public influence was achieved with the foundation of the Anti Corn Law League in 1838, which was headed by Richard Cobden, a Manchester manufacturer who turned out to be a superb strategist and smart tactician on the political way to free trade. On top of this, the ingenious lobbying was favoured by a public climate of discussion that was very receptive to liberal ideas. After all, there were other major moves towards liberalism and away from mercantilism: in 1825 the end of the prohibition for skilled labour to emigrate, which went back to 1719; and in 1842 the end of the prohibition to export machinery, which went back to 1774. Clearly, the British moves towards free trade were part of a general self-confident attitude of an élite in a country that approached a first peak as a pioneer of industrial development. In the same vein, the foundation of the weekly business and finance paper "The Economist" in 1843 can be viewed as an explicit product of this attitude. Quite consistently, this prestigious quality paper, which today has a worldwide circulation, has remained one of the most outspoken intellectual supporters of free trade and Globalization.

In the mid 19th century, Britain was not just another country, but a shining example of a modern society and economy. Thus it is no surprise that Britain's shift to free trade had profound consequences for trade policy all over the continent. With a delay of 1 ½ decades, one country after another went over to a much more liberal trade policy, though not always to fully-fledged free trade. A number of bilateral agreements in the early 1860s set the pace, with the so-called Cobden Chevalier Treaty of January 1860 between Britain and France being of particular political and conceptual importance. Economically, it entailed the abolishment of all French import prohibitions, which were replaced by moderate ad valorem duties of up to, but never more than 30 %. In turn, Britain reduced wine tariffs by a hefty 80 %, allowed the duty-free import of many French products and removed export duties on coal.

Still more importantly, the Cobden Chevalier Treaty was the first to grant most favoured nation treatment (MFN) to the contracting partners. The so-called MFN-clause stipulated that any trade privilege that one party to the contract granted to a third party was to be granted also and automatically to the contractual party. This seemingly innocent principle was to become the basis for the philosophy of non-discrimination that is today a constituent part of virtually all trade agreements and one of the core principles of the multilateral General Agreement on Tariffs and Trade (GATT), which has served as the contractual basis of the World Trade Organisation (WTO).

In the 1860s, further bilateral trade treaties followed: in 1861 between Belgium and France, in 1862 between France and Prussia, and in 1863 between the newly founded nation-state of Italy and various countries. In 1863-6, many other European countries followed suit: Switzer-

land, Sweden, Norway, Spain, the Netherlands, Austria and the few still sovereign Hanse towns concluded agreements. With German unification in 1871, the process turned to its conclusion. Even before unification, Germany had liberalized its internal trade in the Prussian dominated “Zollverein”, which was formally founded by Prussia in 1818, but significantly enlarged in 1834 and 1841. By that time, it included 11 states with almost 87 % of the German population of 1850, thus providing an internal market of considerable size in central Europe. Sure enough, the Zollverein was an instrument of economics and of politics. Economically, it not only facilitated trade per se, but also allowed to proceed smoothly in the construction of an efficient railway system all over the German territories, which made only sense if there were no significant trade barriers anymore. Politically, by excluding Austria, it paved the way towards German unification under the political leadership of Bismarckian Prussia.

By the mid 1870s, Europe had reached a state of remarkably free trade. It has been calculated that, about that time, the average duties on manufacturing products – expressed in terms of ad valorem equivalents – amounted to no more than 9-12 % on the continent, down from a magnitude of about 50 % after the Napoleonic Wars. This was a major achievement, and combined with parallel liberalisations within nations as in freshly united Italy and Germany, it was in fact a heyday of liberal Europe, economically and to some extent politically. After all, one finds quite a few political leaders at that time in the different nations who had either strongly liberal convictions and the necessary political panache or were opportunistic enough to engage in a temporary coalition with liberals:

- in Britain Richard Cobden and Robert Peel, the conservative prime minister who repealed the corn laws against strong opposition in his own party, which split thereafter into a conservative and a liberal wing, the “Peelites”, who later joined the Whigs to form a modern liberal party under the strong leadership of the later prime minister William Gladstone (also a formerly Tory “Peelite”);
- in Germany Chancellor Bismarck, an opportunistic conservative with an ingenious sense for a temporary coalition with the rising parties of liberalism, and Rudolf Delbrück, Bismarck’s “chief of office” who engineered this coalition and, as a staunch liberal without party affiliation, was responsible for trade policy in Prussia and later Germany from the early 1850s until the late 1870s when Bismarck – once again opportunistically – shifted towards protectionism and more state control by nationalizing the railway system;
- in France the emperor Napoleon III, who – though highly authoritarian – was a great admirer of modern liberal Britain and tried to emulate it in terms of an economic performance that he regarded as the basis for the “grandeur” of his own nation;

- in Italy the staunchly liberal Count Cavour, a nobleman from Piedmont who played a major role in Italian unification and gave it a distinctly free-market touch until his premature death in 1861 as the first prime minister of the new nation state Italy, which – after Cavour's death, but in his spirit – concluded liberally-minded trade agreements just as Britain and France had done before.

So much for the history of the path towards free trade. Let us now turn to three major questions that concern the economics and the politics of free trade in the 19th century:

1. Did free trade integrate markets?
2. Why did liberalisation come about at all?
3. What were the distributional consequences of free trade?

We shall try to tentatively answer these questions, thereby once again drawing heavily on the research results presented in the above mentioned book by Kevin O'Rourke and Jeffrey Williamson.

→ 1. **Market Integration**

International trade tends to lead to price convergence of those traded goods that are homogenous. This is so independent of the economic reason for trade, be it lower transport costs and/or lower tariffs. As we know from the historical record, both natural and political trade barriers declined from the first to the third quarter of the 19th century. Hence, for this period, we should observe a price convergence for those commodities that were traded in different places (and can reasonably be regarded as homogenous). In fact we do so.

FIGURES 48-52 show various measures of price convergence for selected commodities, trading places and periods, namely:

- FIGURE 48 the ratio of Bavarian and Prussian prices of wheat, oats, barley and rye for the period 1854-1906,
- FIGURE 49 the price differential of wheat (and meat, respectively) between Britain and the United States from 1870 to 1913,
- FIGURE 50 the price gap between Denmark and Britain for barley and oats from 1870 to 1913,
- FIGURE 51 the price differential of wheat, oats, barley and rye between Denmark and the United States from 1870 to 1913, and
- FIGURE 52 the price differential of wheat, barley and oats between Britain and the United States from 1970 to 1913.

In sum, the emerging picture is one of sustained price convergence in virtually all bilateral cases: the price differential or price gap – how ever defined – tends to become smaller over the relevant period, and so does the amplitude of its fluctuation. There are some minor exceptions to this rule, which are well explainable:

- In the case of Bavaria and Prussia (FIGURE 48), convergence becomes clearly visible not before German political unification in 1871, although economic integration in the form of the “Zollverein” preceded it. This points towards some “unobserved” economic elements linked to political unity such as a more efficient infrastructure or administration that may have supported integration beyond what we can measure in terms of natural and artificial trade barriers.
- The meat price convergence between Britain and the United States (FIGURE 49) begins not before the 1890s, which may be due to the new refrigeration techniques that were not available for transatlantic meat transports before that time.
- Taken together, FIGURES 50-52 show that grain prices converged between Britain and Denmark and between Britain and the United States, but not quite so between Denmark and the United States (notably in the case of barley). The solution to this puzzle may be that, due mostly to declining transport costs, there was a general trend towards (higher) British prices so that, by implicitly filtering out this trend, a juxtaposition of American and Danish prices give a misleading picture of non-convergence.

Note that, for many other commodities or semi-manufactured goods such as cotton, pig iron, iron bars, copper, hides, wool, coal, tin, coffee etc, there is also strong evidence for price convergence. Note also that convergence was not confined to American and Western European markets. It also covered the big grain markets of western and eastern Europe (e.g., Liverpool in Britain and Odessa in Ukraine) and India and East Asia, which were closely linked to the West after the opening of the Suez Canal in 1869.

The obvious conclusion is that the Globalization of markets did work and did matter. However, at this point, it must remain open whether this was due more to a decline in transport costs or to tariff reductions. If anything, the evidence is biased towards declining transport costs as the main driving force. To be sure, the period of tariff reductions ended in the late 1870s, whereas transport costs continued to decline, maybe even at an accelerated speed due to many subsequent innovations in transport technology. And apparently, convergence continued as well, at least between those countries that are considered in FIGURES 49-52 (which are all countries that did not later shift from free trade to protectionism!). Unfortunately, the data basis is very poor for earlier times so that the interplay between transport costs and tariffs cannot be studied more carefully over a longer period of time. As we shall see in

section 3.1.4.2 of this manuscript, the backlash towards protectionism in the late 1870s can best be interpreted as a defensive reaction to the speedy Globalization of commodity markets that was due to some combination of free trade and ever lower transport costs.

→ **2. Reasons for Liberalisation**

There are basically two polar candidates for explaining such fundamental changes of policy regimes as the shift from mercantilism to liberalism in the 19th century: ideas and interests. *Prima facie*, the path to free trade appears to be a classical case for the self-confident contention of John Maynard Keynes in his famous "General Theory of Employment, Interest and Money" of 1936 that the world is ruled by ideas of economists and political philosophers – and little else. After all, Adam Smith and David Ricardo had forcefully argued for the invisible hand of the market mechanism and the gains from trade, and their oeuvres were widely read and appreciated by politicians and the public. On the other hand, vested interests clearly played a part as well: it was a lobby of manufacturers who supported Richard Cobden's Anti Corn Law League, and the lobby of landlords who defended conservative protectionism.

It is likely that the truth lies in the middle, i.e. it was a mixture of new ideas and shifting interests that led to the victory of free trade. This shift can nicely be identified in the personality of the British prime minister Robert Peel who became responsible for the repeal of the corn laws. As a Tory, he started as a natural supporter of the corn laws – not least because he shared that part of the analysis of the classical economists that postulated a negative relationship between corn prices and profits in manufacturing. In essence, the argument said that free trade would lower consumer prices and, in a competitive labour market, also lower the nominal wage proportionally, thus keeping the real wage constant, but boosting manufacturing profits and hence furthering industrial investment and growth. Hence, in real terms, labour would not gain from free trade, despite lower food prices.

However, it gradually became clear that reality was different. In the course of the first steps towards liberalisation in the 1820s and 1830s, nominal and real wages in industry did rise and not fall. In fact studies by modern economic historians have shown that, between 1810 and 1840, the real wage level rose by almost 90 %, i. e. at an annual average compound rate of 2.1 % - after six decades of stagnation or even slight fall from 1750 to 1810. This made Robert Peel confident that free trade was not only a favour done to the manufacturing lobby, but also a measure to support the broad masses of workers.

Graphically, the logic of the situation can be described as in FIGURE 53, which shows an economy of two sectors, agriculture (a), which produces only food, and manufacturing (m), which produces only machines that do not enter consumption. For the sake of argument, we assume labour to be perfectly mobile between the two sectors. The allocation of labour between the sectors is then determined by the intersection of the sectoral labour demand curves, denoted by LD with subscripts a and m, respectively. Note that both LD-curves slope down, though the one for agriculture from left to right (measured from the left-hand side origin) and the one for manufacturing from right to left (measured from the right-hand side origin). On the vertical axis, we measure the nominal wage. Hence labour demand decreases with the nominal wage at a given price level that is assumed to depend strongly on food prices.

In a world with corn laws, the allocation of the total labour force L between the two sectors is determined by the point of intersection A, with a nominal wage of W^* : $0L^*$ labour works in agriculture, $0'L^*$ in manufacturing. A repeal of the corn laws drives down the price of food and thus pushes down the labour demand curve of agriculture. Remembering that, in labour/nominal wage-space, the labour demand curve is the geometric locus of the marginal value product (price times quantity) of labour, the food price decline pushes the labour demand curve of agriculture downwards, in FIGURE 53 by AB. As a consequence, L^*L^{**} of labour migrates from agriculture to manufacturing, which leads to a nominal wage decrease from W^* to W^{**} .

How does the real wage change? In FIGURE 53, it increases because the decline of the nominal wage W^*W^{**} is less than the food price decline AB. However, the extent of the real wage increase obviously depends on the elasticity of labour demand in manufacturing. If this is high in absolute terms, i.e. manufacturing absorbs the labour surplus rather easily, then the real wage increase will be high as the nominal wage falls only slightly. If, however, the elasticity is low, then the real wage increase will be small. In the extreme case of an inelastic labour demand in manufacturing, there will be no real wage increase at all, and also no structural change from agriculture to manufacturing.

How was reality perceived by politics in Britain in the second quarter of the 19th century? Apparently, Robert Peel more and more believed to see the elastic case, and not the inelastic one. Hence he – and many others – took the repeal of the corn laws as politics in favour of the broad masses, and not just a manufacturing lobby. This appeared all the more so because, at that time, a large part of any worker's expenditure was for food, and little was left for other goods. Even then, however, Robert Peel did something that was clearly not in the

interest of the class that supported his own Tory party, which in fact later split over the issue as mentioned above. Hence neither manufacturing nor agricultural interests alone can explain the shift. The power of ideas and the power of interests apparently came together.

It is an irony of history that, after the repeal of the corn laws, it took more than two decades until food prices did in fact fall substantially. This led some traditional historians to the conclusion that, economically, the repeal did not really change much. This conclusion is obviously flawed. To evaluate the economic effect of the corn laws and their repeal, one needs a carefully worked out counterfactual scenario of what would have happened in the absence of repeal. That requires a (computable) general equilibrium model to assess. In such a model, it is – among many other things – the rising demand for food due to the rapid increase of world population that has to be taken into account. If this is properly done (and it was done econometrically), then the unambiguous result is that, for most of the 30 years after repeal, the counterfactual no-repeal price of food would have been more than 20 % higher than the actual price.

→ 3. Distributional Consequences

It remains the question of how the distribution of income was affected by the repeal of the corn laws. To be sure, the lobbies at that time had a clear conception of the gains and losses: landlords were to lose, manufacturers were to gain, and labour – as described above – was to gain more or less. Were these conjectures borne out by reality? The Irish economic historian Steven O'Rourke has tried to answer this question empirically – on the basis of a computable general equilibrium model, which consists of five economic sectors, namely two agricultural ones: tillage and pasture; and three non-agricultural ones: mining, manufacturing and services. He asked what would have happened if corn laws had been repealed earlier, say, not in 1846, but in the early 1820s.

Without going into the details of his analysis, the results can be summarized as follows:

- Assuming that Britain had no influence on world market prices (“small-country assumption”), there would have been a sharp decline in agricultural employment (by more than 20 %), a slight decline of the nominal wage of labour, but a hefty decline of food prices by 25 %, together implying a real wage increase of about 23 %. Grain rents would have fallen by almost 50 %, non-grain rents by almost 7 %, but manufacturing profits would have risen by a hefty 25 %.

- Allowing Britain being an economically large country with an influence on world market prices, the repeal of the corn laws has a cost to Britain in the form of a terms-of-trade loss that is induced by her rationing of import demand and her increasing of export supply. Under defensible assumptions, a model with terms-of-trade losses yields a somewhat lower decline of grain rents by 42 % (down from 50 %), a real wage increase of 12 % (down from 23 %) and, most remarkably, a decline of manufacturing profits by 22 % (in contrast to the gain of 25 % estimated without the terms-of-trade effects).

All in all, these results show that the vested interests (and also the Tory prime minister Robert Peel) shared a quite realistic assessment of the gains and losses of the corn laws' repeal. Landlords resisted it, and they were right to do so from the standpoint of their group interest. Labour – to the extent that it had a political voice – favoured repeal, and for good reason as liberalisation made the real wage rise. Only the outcome for manufacturing interests remains ambiguous due to the working (or not) of terms-of-trade effects. Sure enough, there is a more general issue behind the net balance of manufacturing. Economically, the analysis shows that the location of the corn laws' burden depends on whether Britain was a "small" or a "large" country in the sense of trade theory: if it was small, the burden was carried by its own manufacturing sector; if it was large, the rest of the world paid the bill.

What was the truth? That certainly depends on the time horizon we consider. In the short run, it is realistic to assume that Britain was "large" – as the dominating industrial power in the second quarter of the 19th century. In the longer run, Britain is better assumed to be "small" – with other continental and overseas countries, notably Germany and the United States, quickly catching up as major sources of manufacturing supply and food demand. Hence, in the long run, it is highly unlikely that Britain could have gained from maintaining the corn laws in place. And, in its lobbying against the corn laws, British manufacturers may have hurt themselves in the short run, but certainly not their industry in the long run.

Note that the negative terms-of-trade-effects have been a big and controversial topic among Anglo-American economic historians. One of them, Donald McCloskey, drew the conclusion that, in welfare terms, Britain hurt itself by the repeal of the corn laws because it incurred a welfare loss of 2 % of GDP according to his own calculations. Once leaving the narrow world of general equilibrium modelling, this conclusion can hardly be accepted: not in the long run, as the price elasticity of supply has a tendency to increase over time as was certainly true in 19th century Europe; but also from a standpoint of the political economy of international trade

because liberalisation in Britain led others to follow suit and thus ushered in a period of free trade, with massive gains from specialisation for all nations, including Britain.

In addition, one must not forget that the corn laws' repeal accelerated Britain's structural change away from agriculture towards manufacturing, i. e. to that sector of the economy, which was to realize by far the fastest technical advances and productivity increases in the decades to come. From the point of view of modern growth theory with its emphasis on endogenous progress of technology, this was likely to be a big gain for the British economy. Note also that the repeal induced a structural change within (shrinking) agriculture – at the expense of tillage, which produced tradable commodities, and to the advantage of animal husbandry and dairy farming, which tended to produce non-tradable perishable goods like milk or partly tradable ones like butter and cheese. It is this structural change within agriculture that may also explain why Britain did not return to protectionism when others did from the 1970s on. Apparently, the lobby of "tillage agriculture" was by that time too weak to have a strong political influence, even in the Tory party of the later prime minister Benjamin Disraeli, who in the 1840s was still strictly against the repeal of the corn laws.

Note also that Britain's move to free trade had important consequences for Ireland, which – under British rule – did not have an independent trade policy. Being a region with virtually no industry, the repeal of the corn laws induced a structural change within agriculture as it did in England: less tillage and more animal husbandry and dairy farming. However, it may also have led to mass emigration. That is at least how it appears when considering that the repeal of the corn laws in 1846 was followed by the tragic Irish famine of 1846-8, with an altogether unusual three-year sequence of awfully bad potato harvests. As no Irish industry stood ready to provide jobs, people were in a sense pushed out of the country by hunger – to the industrial centres in Britain and the eastern United States. However, one must be careful in identifying causes and effects correctly: taking a longer run view well beyond the immediate consequence of the famine, it is clear that, to explain Irish mass migration, the "push factor" of liberalisation was probably much less important than the "pull factor" of rising wages in Britain and the United States. In fact, between 1800 and the 1870s, real wages in Ireland rose by almost 30 %, and they mostly did so as a consequence of the labour scarcity across the Irish Sea and the Atlantic Ocean. Hence, contrary to popular myths, free trade by itself can hardly be made responsible for the typically Irish development in the 19th century. We shall return to these matters when dealing with international migration in section 3.1.6 of this manuscript.

3.1.4.2 Protectionist Tendencies 1878-1914

The honeymoon of free trade in Europe was soon over. From the late 1870s on, one continental country after another introduced tariffs, both on agricultural and manufacturing products. For a start, the tariffs turned out relatively moderate, as becomes clear if one expresses them (as we do) in terms of their ad valorem equivalent, i. e. as a share of the product value. Subsequently, however, they were raised to fairly substantial levels.

E. g., in 1879, Germany introduced specific import tariffs on wheat, rye and other cereals of 6-8 % (in ad valorem terms), and subsequently raised them to levels of 33 % on wheat and 47 % on rye in 1885 and 1887. After a brief reduction in the early 1890s, duties were further increased in 1902 (with the so-called “Bülow-tariffs”). In France, the shift to protection began in the mid 1880s and reached a peak with the so-called Meline-tariffs in the early 1890s, which levied specific duties of at least 10 % (ad valorem) for general agricultural products, but topped by wheat with a duty of 32 %. There were other countries that also joined the protectionist camp: Italy, Spain and Portugal took action in the late 1870s; Sweden turned protectionist in the 1880s, roughly parallel to Germany. However, not all countries followed suit. In particular, the United Kingdom, the Netherlands and Denmark as well as Belgium and Switzerland largely remained free traders, at least in agriculture. Thus, on the eve of World War I, Europe had become a divided continent in terms of trade policy.

FIGURE 54 shows various measures of protectionism in manufacturing and agriculture for those twelve countries for which appropriate data have been analyzed by economic historians. All in all, by 1913, there are three groups of countries:

- the overall free traders with no or low tariffs in agriculture and in manufacturing: the United Kingdom (including Ireland) and the Netherlands;
- those with free trade in agriculture, but more or less protectionism in manufacturing: Belgium, Denmark, Norway and Switzerland;
- the overall protectionists with high tariffs in agriculture and considerable protection in manufacturing: Germany and Sweden as well as France, Italy, Portugal and Spain.

The grouping of countries gives some hints at where the political motives for tariff protection lay. Those countries with an advanced manufacturing industry and a relatively small and/or efficient agriculture tended to overall free trade or to free trade in agriculture and some protection for industry; and those who had a large and less efficient agriculture based mainly on grain production (and not animal husbandry and dairy farming!) tended to overall protectionism. Apparently, the exogenous shocks that the Globalization of markets entailed affected

the various countries in different ways so that different degrees and structures of protectionism (or none at all) came about.

In most places, the search for the appropriate trade policy stance was accompanied by big political controversies between free-traders and protectionists. Even in the United Kingdom, the country with the most deeply rooted free-trade tradition, the early years of the 20th century brought heated public debates between liberals and conservatives about whether or not to join the protectionist camp. Only after a sweeping political victory of the liberals in the parliamentary elections of 1906 did this debate end with a re-affirmation of the British free trade position.

It is Germany that delivers a particularly good example to demonstrate the political economy of the new protectionism. By the mid 1870s, the grain production in the eastern parts of the country came under massive competitive pressures once costs for long-distance transport had tumbled and free trade was firmly established. As the agrarian lobbies had influential spokesmen in the governments of Prussia and the newly founded Empire, the pressure for political action mounted. The breakthrough came once an alliance was in sight between western industrialists of the highly cartelized iron and steel industries and eastern agrarians – against the interests of (grain-consuming) labour and of highly export-oriented and innovative manufacturing branches like chemicals and electrical engineering, which were fast-growing, but still relatively small. Chancellor Bismarck, a conservative agrarian by birth and upraising and always keen to secure his power base, stood ready to shift gears towards protectionism and to support this new alliance, which was aptly called the marriage of “iron and rye”. Economically, this meant high tariffs in agriculture and moderate tariffs in manufacturing (see FIGURE 54). Politically, the long-run consequences were even much more fundamental: the alliance led to a massive political weakening of German liberalism and thus – internally – to a political shift towards a paternalist conservatism and – internationally – to a worsening of Anglo-German relations, which had always rested on the political inspiration that German liberals drew from the successful British example.

To be sure, the Bismarckian shift from liberalism to protectionism ushered in a lively and long-standing debate among German intellectuals, which left deep traces in the literature, which are remarkably topical for today’s Globalization debate. Liberals like Max Weber, the academic doyen of German sociology, were typically not only defenders of commerce and trade, but indirectly of all those dramatic transformations that came with the structural change of industrialization and Globalization. In a sense, they took the part of those who today argue that capitalist Globalization – if accompanied by proper politics – is a development in an inev-

itable and essentially welcome direction, and a good opportunity to come closer to a world of widespread prosperity. On the other side stood those intellectuals like, e. g., the prominent public sector economist (in traditional German: "Staatswissenschaftler") Adolf Wagner who coupled the case against free trade and commerce with a kind of conservative-romantic condemnation of contemporaneous urbanisation. In the same spirit as today's critics of Globalization, they argued that the growth of cities and the massive migration from the land to the towns went much too far and should be slowed down by appropriate policies.

Did the new protectionism succeed in its aim to inhibit world market integration for the protected goods and production lines? Some traditional historians have argued that it did not because the decline of transport costs was such an irresistible force that it dwarfed all efforts at protection. As modern research has shown, this view – if taken literally – is untenable. Economically, tariffs did matter: FIGURE 55 shows that the price gap for wheat between some protected and non-protected markets (France, Bavaria and Sweden vs. Britain and Prussia vs. Denmark) widened substantially in the 1880s and 1890s from close to zero up to 25-50 % where it stayed up to the eve of World War I. This is a clear piece of evidence for the economic power of the tariffs. In the same vein, international price spreads for a variety of grains developed very differently among free trade nations and among protectionist ones or between free traders and protectionists (FIGURE 56): among free trade nations, they shrank, but in all other cases, they remained as they were or even increased.

Despite this evidence, however, traditional historians may be conceded to have a point if one puts the late 19th-century backlash towards protectionism into a wider perspective. After all, high tariffs for agriculture and generally moderate tariffs for manufactures did not prevent Globalization from continuing. Clearly, they did lead to a retardation of structural change in the protectionist countries, and in some of them like Germany, they tilted the political balance in favour of an anti-liberal stance that probably incurred large costs in the longer run. However, the modernisation did go on at a speed that was high by all standards available to contemporary observers and still to us today.

Take, e. g., Germany: on the eve of World War I, Germany had 36 % of its labour force in industry, thus being one of the most industrialized economies of the world – and that after 30 years of agrarian protectionism. And the fastest growing industries were not protectionist iron and steel, but export-oriented chemicals and electrical engineering. To be sure, the share of agricultural employment was still more than one third, with a large part of it in protected grain production, and this had certainly to do with high tariffs on rye, wheat and other cereals. In view of the fact, however, that there was a massive migration from the agrarian East to the

industrial western and central regions of the country, it is hard to speak of a severe backlash in terms of structural change. More to the point, the wave of protectionism may be described as a kind of “leaning against the wind” of a Globalization that was fuelled by new industries, new technologies and a rapid decline of transport and communication costs. Thus re-interpreted, traditional historians may not be so much off the truth.

A similar point can be made for protectionism in the New World. As early as 1791, Alexander Hamilton had made a case for protection in the United States. In his famous “Report on Manufacturing”, Hamilton argued in favour of an infant industry protection of the fledgling American manufacturing industries, which should grow into (internal and external) economies of scale behind tariff walls. For a while, the mighty British manufacturers should be kept away from the American home market until a level playing field was established after American industry had learnt how to produce with higher productivity and at lower cost.

During the Napoleonic wars, the United States made a first venture into protectionism with the so-called Embargo Act of 1807. However, this seemed to remain an episode as it was repealed two years later. With manufacturing growing notably in the North-East, the 1820s brought two further important protectionist laws, culminating in the “Act of Abominations” of 1828, which was also strongly supported by the German economist Friedrich List, who at that time lived in the United States and – still today – is referred to as one of the intellectual fathers of the Infant Industry Argument in Trade Theory. To be sure, the American move towards the protection of manufacturing led to major internal tensions within the United States, as the free trade southern states massively opposed any conflict with Britain, their major export market for raw cotton. Thus, in the Compromise Act of 1833, tariffs were somewhat reduced, but the “tariff question” remained one of the major issues of political conflict between North and South up to the Civil War.

In the New World, the United States did not stand alone with their approach to protection. In the later 19th century, Canada, Australia and also Latin American countries went over to some sort of protection for manufacturing, once the new industries began to grow in their countries. In some cases, the tariffs that were eventually reached were quite high by any standard. E. g., the so-called Victorian tariff of 1865 of 10 % (ad valorem equivalent) was successively raised up to 45 % by 1893, which amounts to a degree of manufacturing protection that was not reached in any place of Europe, not even in the Mediterranean countries.

All in all, the picture of trade policies in different countries makes clear that the world of the late 19th and early 20th century was anything but a “free-trade paradise”, as friends of Global-

ization might be tempted to call it, or a playground for unfettered market capitalism, as foes of Globalization might think. Rather it was a world in which manifold vested interests pushed through substantial deviations from free trade and commerce. On the other hand, it was also a world where the sheer power of technical progress, growth and structural change overran the conservative forces of retardation, at least until World War I. In this respect, it resembles most recent times, and it is not surprising that the policy debates of the late 19th century have so much in common with the Globalization debate today. We shall return to these matters in section 4 of this lecture.

3.1.5 International Capital Movements

International capital movements reached a peak in the period 1870-1914. This can be seen from FIGURE 57, which shows foreign investment as a percentage of domestic savings for three large European countries: the United Kingdom, Germany and France. The numbers for five-year intervals fluctuate strongly, but yield an average, which is by all means impressively high: from 1870 to 1913, roughly one third of all British savings were invested abroad. For Germany, the respective share is 12 %, and for France 16 % - all levels that are much higher than those reached most recently, i. e. at a time of allegedly globalised international capital markets.

FIGURE 58 provides a long-term comparative view of the extent of capital flows relative to general economic activity – measured as the absolute value of the current account as a percentage of GDP. The numbers are five-year averages for a sample of 14 countries for 1870 to 1990, among them all the major industrialized countries of today. They reveal a clear time profile: very high levels in the range of 3 to 5 % between 1870 and 1913, a marked decline in the interwar period down to a low of about 1 % by the 1930s and a gradual increase not before the mid 1970s from 1 % to 2.5-3 % by 1990. Hence, not before the 1990s, did the extent of international capital flows reach again the dimension that prevailed before World War I.

There is a straight econometric way to evaluate the degree of internationalization of capital markets over time. It goes back to Martin Feldstein and Charles Horioka who argued that a straight regression of a country's investment share in GDP on its savings share yields a measure of the degree of international capital market integration:

- if the regression yields an estimate of the slope coefficient β that is close to one, there is little or no international capital mobility as, apparently, countries finance their own investment through their own savings;

- if the regression yields an estimate of the slope coefficient β that is significantly smaller than one and/or a correlation between savings and investment shares that is very low, then the reverse holds, i. e. there is a large lending and borrowing across borders, leading to a loose link between savings and investment shares.

FIGURE 59 shows the results of such a cross-country regression analysis for a long sequence of five-year-periods from the early 1870s to the early 1980s. The message of the graph is unambiguous: low estimates of β ($\beta \approx 0,5$) for the time before World War I, a sharp increase of β in the interwar decades and a high β ($\beta \approx 1$) for the period after World War II, until a substantial gradual decline sets in during the early 1970s, which leads down to a low β ($\beta \approx 0,6$) by the 1980s. This econometric evidence nicely complements the more descriptive picture presented above. Sure enough, the pre-World War I-period was a time of high capital market integration, even by the tough standards of modern times.

Note that other studies of economic historians strongly support the view that financial markets were well integrated in the late 19th and early 20th century. In particular, there is ample evidence that nominal interest differentials between the major financial centres across the Atlantic (e.g., London and New York) were relatively small in the second half of the 19th century. And they became ever smaller once the modern communication technologies led to a quite dramatic speeding up of news transmission over long distances (see section 3.1.3).

So much for the forces of financial market integration, which were indeed impressive. Nevertheless, it is important to keep in mind what precisely the many pieces of evidence for integration indicate and what not. Clearly, they do indicate that, by today's standards, there were remarkably large and sustained imbalances of saving and investment internationally, which could be sustained in well-developed international financial markets. There were lenders in one country and borrowers in another, and the **net** balance of the international lending and borrowing was large. However, the evidence does not imply, that the **gross** flows of capital were anything as large as they are nowadays. This means that, today, there is a vastly bigger portfolio diversification in all countries than was the case before World War I. But much of this diversification – say, Americans holding European shares and bonds and Europeans holding American shares and bonds – cancels out in a net macroeconomic balance that focuses on the current and capital account.

Also, big surpluses and deficits in the macroeconomic balance say nothing about the extent of **direct** foreign investment (as distinguished from **portfolio** investment). Today, there are certainly many more companies that strategically invest in other countries for reasons of

business prospects, resource or factor costs and/or market penetration, which has not much to do with the straight rate of return motivation that underlies portfolio investment. We shall return to these matters when dealing with the more recent times of Globalization in parts II-VI of the lecture.

FIGURE 60 shows the major countries and regions of destination of British foreign investment in the years preceding World War I. The numbers indicate that more than two thirds of all investment went into the New World British empire, the United States and Latin America – all “frontier” regions, which were rich in land and resources. These regions attracted not only capital, but also labour. Hence the picture that emerges for Britain (and also for other capital-exporting countries) is one of a parallel movement of all mobile factors of production – labour and capital – towards the immobile ones: land and natural resources. This picture receives full support when taking into account the types of foreign investment that prevailed: most of the British foreign investment went into bonds and to a lesser extent shares that were issued in London financial markets by foreign governments and firms to finance infrastructure investments in railways, ports, mines or other projects to allow the economic development of the frontier.

Note that, as such, the capital flows were not flows from rich to poor countries in any meaningful economic sense. To be sure, Britain was more industrialized and had a higher per capita income than Canada, the United States, Argentina, Chile or Australia. But real wages tended to be higher in these frontier regions than in Britain and certainly higher than in the rest of Europe. Hence lending capital and thus “chasing” emigrant labour in the New World was by itself no contribution to an international convergence of real wages, and maybe also not a contribution to a convergence of living standards in general. We shall return to this important issue below when discussing the causes and consequences of mass migration in the 19th century.

What made financial capital so mobile at that time? Clearly, technological improvements played a major part. Beyond that, however, it is tempting to search for those institutional conditions that may have been particularly favourable for capital market integration. Economic historians agree that, in this respect, a prime candidate is the exchange rate regime, the Gold Standard, and the role that sole world financial centre of the time. To these we now turn.

3.1.5.1 The Gold Standard

The period from about 1870 up to the eve of World War I was not only a time of a very stable system of fixed exchange rates between national currencies, the first that has ever existed over several decades: the Gold Standard. What was it and how did it work?

In essence, the Gold Standard consisted of two major commitments, which were made by governments and central banks:

- The national money was defined in terms of a given quantity of gold, i. e. there was a fixed price of gold in national currency, and vice versa.
- The central bank was obliged to buy and sell gold at this very price, i. e. there was free convertibility between the national money and gold.

Together, these two commitments had two major consequences:

- Exchange-rates between gold-backed currencies were fixed, due to the ever present option of private gold arbitrage between countries.
- The national money supply depended on the gold reserves of the central bank, which were controlled by market forces. Hence there was no autonomous monetary policy.

In a sense, the Gold Standard was the textbook case of a fixed exchange rate, although it was never deliberately designed as such. In fact, historically, it developed in a kind of evolutionary process over a very long period of time when central banking began to develop as a form of market domination of private banks that issued currency and were then chartered as monopolistic providers of a national money. The Bank of England was the first to adopt a full legal gold standard in 1816, but it took more than half a century until other emerging central banks joined the “gold club”. For a long time they stuck to some sort of bimetallic standard because silver was in much more plenty worldwide supply. This changed with soaring gold production overseas in the middle of the century (see FIGURE 61) so that ever more nations went over to gold backing. The final breakthrough came with the conversion of newly united Germany to the gold standard after the German-French war, which delivered a rich yield in terms of redemption payments in the form of precious metals for the German government and the newly founded central bank (the “Reichsbank”). De facto, the United States followed suit a few years later, though the official American adaptation of the Gold Standard did not take place until 1900.

The fundamental mechanism that underlays the Gold Standard was aptly described in the 18th century by David Hume, the famous philosopher of empiricism, who thus became the first international macroeconomist of the world. “Price-specie-flow-mechanism”, this is how he called it, and it runs like this:

- A deficit in the balance of payment – say, a deficit in the trade balance, caused by more imports than exports – leads traders to have an excess demand for the foreign currency, and an excess supply of the domestic one. The excess demand and supplies will be met by buying gold for domestic currency at the fixed gold price at the domestic central bank, thus reducing its gold reserves, and selling this gold for foreign currency at the foreign central bank, thus increasing its gold reserves.
- The gold flow between central banks finds its counterpart in a change of the money supplies: the domestic money supply contracts (as the central bank has to buy domestic currency for gold), and the foreign one expands (for an analogous reason), thus leading to a deflationary price trend in the home country and an inflationary one in the foreign country. This will tend to make domestic goods cheaper and foreign goods more expensive. And this pulls towards a new equilibrium of the balance of payments between the countries.

Clearly, the price-specie-flow-mechanism makes an autonomous monetary policy impossible, at least beyond a certain narrow limit, which is determined by the leeway to “sterilize” gold and currency flows in and out of the central banks’ reserves: to the extent that the bank is able counteract these flows by appropriate transactions that keep the money supply constant, the bank may preserve a certain degree of policy freedom. However, this degree will be minimal whenever there is what may be called a fundamental disequilibrium – meaning that the excess demand or supply of gold is here to stay and can in fact only be corrected by a corresponding money supply change with its “automatic” price level effects that feed back into goods markets in an equilibrating fashion.

Of course, transitory disequilibria may be overcome by lending and borrowing in international markets, and this is exactly what happened in the actual practice of the Gold Standard. Given the persistently high current account surplus of the United Kingdom, the London financial markets stood ready to provide credit for those countries that ran into current account deficits. Apparently, this worked relatively smoothly: at no time in the almost 4 ½ decades of its existence did the Gold Standard approach a state of massive realignment needs and acute emergency – as did the Bretton-Woods system in the early 1970s. Some economic historians argue that the system worked asymmetrically in the sense that the pressure to contract in case of a deficit was stronger than the pressure to expand in case of a surplus of similar magnitude. This may be true, but even if it is, the extent of the asymmetry was never large enough to destabilize the system, although it might have introduced a suboptimal deflation-

ary bias. In any case, it took as much as a whole-scale World War to put an end to gold convertibility and thus to destroy the major pillar on which the system was built.

Given this apparently quite good – or at least not so bad – record in terms of automatic international stabilization, the question remains how the Gold Standard fared in terms of price stability. After all, the mechanics of the system made price stability heavily depend on the supply of gold relative to goods: if the gold supply were to increase fast, the money supply would have to track this increase as gold would be sold to the central banks at the constant parities to obtain domestic and foreign currency, and this might lead to inflation when, in the end, too much money was chasing not enough goods. Conversely, a drying up of gold supplies might lead to sustained deflationary pressures all round.

To be sure, the gold supply grew substantially over the 19th century due to a gold production that grew in trend, but also fluctuating in long waves and swings (see again FIGURE 61). And these long-term fluctuations of the production appear to have had a substantial effect on price inflation as FIGURE 62 shows. In fact, the 19th and early 20th century can be roughly divided into four periods of price regimes, two of secular price declines (1816-49 and 1873-96), and two of secular price rises (1849-73 and 1896-1913). And the periods of rise and decline do correspond to likewise long-term swings in gold production, which by themselves correspond with long-term swings of economic activity as measured by measures of gross domestic product. Given the mechanisms of the Gold Standard, it is not surprising that the long-term price trends – and also the more short-term price fluctuations – were highly synchronized across nations, as the wholesale price statistics in FIGURE 63 for the United Kingdom, Germany, France and the United States nicely indicate. In the core period of the Gold Standard, which covers the deflationary regime of 1873-96 and the inflationary one of 1896-1913, the aggregate price changes over the period are remarkably similar for all four countries, ranging from -53 % to -40 % in 1873-96 and from +39 % to +56 % in 1896-1913.

So much for the facts of inflation. Does this overall picture qualify the Gold Standard as a success story or a failure in terms of price stability? The answer depends on the standard we use. If the standard means price stability in the strict sense of no price inflation (or deflation) over any longer period, then obviously the Gold Standard failed. However, viewed over the very long period 1816-1913 or the core Gold Standard period from the early 1870s to the eve of World War I, the price level remained stable – or, in more precise statistical terminology, stationary. And this experience differs sharply from the inflationary trends that prevailed for a long time after World War II up to the present, where a general price decline was out of the question at any time. On top of this, the average compound rates of price inflation even in the

boom periods were low by later 20th century standards: e. g., from 1896 to 1913, average price inflation ranged from 2.7 % p. a. in the United States over 2.2 % p. a. in France and Germany down to 2.0 % in the United Kingdom, which is all much less than the inflation rates that prevailed from, e. g., the 1950s up to the 1980s. Hence, even in boom times, the loss of the value of money in terms of goods was relatively moderate, and a spirit of stability survived all throughout the period.

This can also be inferred from the development of long-term nominal and real long-term interest rates, the latter being defined as the nominal rate minus the rate of inflation as measured by the actual percentage change of the wholesale price index. FIGURE 64 shows these variables for the United States for the period 1870-1913. The graphs reveal that the nominal rate gradually declined from a high of 7-8 % in the early 1870s down to 4-5 % in the early 1890s where it stayed up to the eve of World War I. In turn, due to the much stronger swing of the price index first down and then up, the real interest rate declined from a high of about 10 % in the 1870s down to a low of -1 % in 1900; from there, it gradually rose again to about 4 % by 1913. Apparently, the nominal rate did not really react strongly to inflation in the way the so-called Fisher equation predicts, which says that, once inflation is anticipated in financial markets, the nominal rate rises to keep the real rate constant, as lenders and borrowers make intertemporal calculations in terms of real and not of nominal magnitudes. Apparently, the financial markets in the late 19th and early 20th century did not anticipate the inflation and continued to charge unchanged nominal rates. Economically, this may be interpreted as a form of myopia, but it may also be regarded as a powerful indicator for the fact that the “price climate” was regarded as essentially stable, despite the apparent inflation of 2-3 % p. a. in boom times. Hence, if anything, the financial markets had quite a bit of trust and confidence in the stability properties of the monetary framework. Albeit implicitly, gave a good grade to the Gold Standard.

3.1.5.2 The Role of London

The Gold Standard was certainly more than an abstract mechanism that ruled the world of gold and currencies. In the perception of the markets and the public, it was closely linked to the City of London as the absolutely dominating world financial centre in the late 19th and early 20th century. Clearly, in the “thick” London capital markets, all modern financial instruments were available, and new ones were invented at a rapid rate. This provided the ideal ground for any kind of international refinancing that was required in trade, commerce and investment. And it also provided the solid commercial basis that a fixed exchange rate system needs to operate smoothly.

To be sure, the reputation of London was strongly enhanced by the fact that it was located in Britain, as the capital of a nation that deserved to be called a “benevolent hegemon” of world financial markets at that time. This was so for essentially three reasons:

- Britain had a most comfortable external position. Consistently, it ran high current account surpluses all over the period 1870-1913. In macroeconomic terms, this meant that the country had to accept a relatively modest internal absorption, avoiding extreme booms in domestic consumption and investment. Apparently, the British public was ready to accept this “sacrifice”, which is the typical attitude of a nation that naturally takes a leading role in the world economy and in international politics.
- Britain was widely regarded as the champion of international economic relations, and of the liberal rules that largely governed trade and commerce between nations. It was therefore not only of economic, but also of great symbolic importance that the United Kingdom never relapsed to protectionism, though there were considerable pressures to do so within the country. This strengthened the political credibility of a liberal economic order of which the Gold Standard was an integral part.
- Britain was at the peak of its political and economic power by 1850-70, and this power – though slowly fading – was still big enough in the following decades to give it a strong political voice that could not be ignored by anybody. Combined with the typically British idea of a balance of power, this gave the international political order a certain stability, which also supported the working of the Gold Standard and the City of London as the market where the hegemon supported or provided whatever “lending of last resort” was necessary to keep the system working.

To be sure, all these were very specific conditions, and there was no certainty at all that they would prevail for ever. In fact, World War I destroyed them. And as we shall see in section 3.2 of this manuscript, all efforts to restore them failed. Not before the second half of the 20th century did anything comparable emerge in the form of the Bretton Woods system of fixed exchange rates and the United States becoming the new hegemon. However, as we shall see in section 3.3 of this manuscript, this new constellation did not survive for longer than 25 years, a much shorter life cycle than the one of the Gold Standard. Since 1973, we have floating exchange rates between the major currencies of the international financial markets. Only within Europe a system of fixed, but adjustable exchange rates was re-established in

the late 1970s. It laid the foundation of the European Monetary Union with the Euro replacing the national currencies in large parts of the European Union.

Note finally that one should beware of idealizing the time of the Gold Standard as a happy “Golden Age” of stability, growth and free trade. To be sure, economic success does not necessarily imply political desirability as it may depend on conditions that, from today’s vantage point, appear to be unacceptable. E.g., the time of the Gold Standard was also the period of what political scientists call “imperialism” or “colonialism”, and Britain’s political strength depended not least on its imperial power base. Thus, historically, the success of the Gold Standard and of all that came with it might be regarded as one of the consequences of imperialism, i. e. of a particular political power structure with a dubious reputation.

At this point, we can leave this issue aside because most modern economic historians agree that imperialism was of much less importance for the economic development of the 19th century than many Marxist analysts once suspected. After all, many countries – those with colonies and those without colonies – did industrialize, grow and integrate into world markets, and there is no empirical link between a country’s economic performance and a country’s degree of colonial imperialism. E. g., in the period 1870-1913, Germany, Austria-Hungary and the United States, three countries without any significant colonial base, caught up rapidly to the colony-rich United Kingdom. In turn, the long-standing colonial powers France, the Netherlands and most of all the United Kingdom grew disproportionately slow. Hence, quite obviously, colonialism does not deliver an explanatory key to the major economic trends of the time up to World War I. Nevertheless, from a broader historical standpoint, imperialism may deserve consideration as one of the political pillars of the “Pax Britannica” that prevailed for so long in that period. However, this kind of consideration goes well beyond the limits of this lecture, which is above all on economic and not on political matters.

3.1.6 International Migration

“The first era of integration: 1820-1914” – this is the title of section 3.1 of this manuscript. As we have seen, it was a period of fast and widespread industrialization, of a massive expansion of international trade and commerce, and of a growing integration of financial markets. Besides all this, it was also the time of massive international migration. In many respects, the second half of the period from about 1870 to the eve of World War I marks a peak of international labour movements notably across the Atlantic, from Europe to the New World in America, both North and South.

FIGURE 65 summarizes the quantitative picture of this migration. It shows migration rates by decade (per 1000 mean population) for 15 European and five American countries from 1851 to 1910. The basic message of the table is simple, both over time and across countries. Over time, the flow of transatlantic migration reached a peak in the 1880s. Thereafter, emigration declined in some countries (like Germany, Britain and Ireland), but rose in others (like Italy, Portugal, Spain and Austria-Hungary). Across European countries, emigration rates differed strongly, with the highest in Ireland, Norway and Britain and the lowest in France, Belgium and the Netherlands. Note that, in terms of absolute numbers, large countries like Germany and Austria-Hungary were major suppliers of migrants although their emigration rates were never exceptionally high.

In standard textbook economic models of international migration, it is the real wage difference between the home and the receiving country that determines migration decisions. Other things being equal, labour migrates to the place where its marginal product and thus its real wage is highest, whatever the reason for this may be (e.g. a higher level of technology or an abundant land or capital supply). In the 19th century, the countries with the highest real wage rates were typically those with abundant land and scarce labour, and these were also the classical immigration countries like the United States, Canada or Argentina. *Prima facie*, this supports the theoretical idea of the wage gap.

In view of the prominent theoretical role of the wage gap, it is tempting to expect a positive correlation of the “real wage gap” between the receiving and the home country with the rate of emigration. In FIGURE 66, the variable denoted by “B” shows appropriately calculated wage gaps, measured as the wage ratio of the home to the receiving country (with the wage being defined as the weekly wage in urban areas): the lower this ratio, the higher the wage gap in the respective home country relative to the receiving country in the decade under consideration. Note that, in FIGURE 66, the variable “A” denotes the real wage rate in the home country as a percentage of the British level of 1905, thus delivering a measure of the wage gap relative to the industrial leader of Europe with the highest wage level.

The message of FIGURE 66 is unambiguous: the wage gap by itself is not able to explain the vast variation of migration rates over time and between countries. Apparently, other factors played a major role as well. Three factors stand out in importance:

- The cost of migration may be so high that, below a certain threshold level of income, no migration is forthcoming because potential migrants are simply unable to cover the expenses for travelling to and settling in a new place. Up to this threshold level, mi-

gration is supply-constrained and labour finds itself in a kind of geographical poverty trap. Only above the threshold migration is it demand-constrained and can thus react at all to wage gaps. All this is particularly important for transatlantic migration, which was very expensive to begin with, but in the course of the 19th century became much cheaper with the dramatic decline of transatlantic transport costs.

- Industrialization and urbanization of an economy tend to raise not only the degree of internal mobility from agriculture to industry and from the countryside to the towns, but also the mobility between nations. Once people start to live in cities, they lose their deep rural roots and may thus regard a transatlantic migration as a natural option. Hence the elasticity of emigration with respect to a wage gap may increase substantially with urbanisation, which may loosen or even reverse the link between a wage gap and emigration flows. After all, more industrialized countries tend to have a higher real wage level than rural ones.
- There is much anecdotic evidence of “chain migration”, which means that, once a kind of community of relatives, friends and colleagues from a home nation has developed in the receiving country, the cost of further emigration for those who have remained in the home nation is substantially reduced. The reason lies in the typical services that this community delivers to later migrants – in the form of remittances, pre-paid tickets of travel, free room and board upon arrival and also contacts to potential employers. If chain migration is real, it typically shows up as a path dependence in the data (or, in econometric models, as a strong and significant autocorrelation and/or a strong explanatory power of the stock of immigrants in the receiving country).

In an elaborate econometric study on 19th century migration streams, the two economic historians Timothy Hatton and Jeffrey Williamson tested some of these (and other) explanations for the peculiar patterns of transatlantic migration. Without going into any details, the general message of their estimates can be summarized by a stylized picture of the “factors in the emigration cycle”. These are presented in FIGURE 67, which shows the contribution to the trend of four factors over a period of eight decades as they follow from the econometric estimates. The graphs make clear that other explanatory factors than the wage gap were indeed of importance:

- population pressure due to demographic factors, which peaked in the sixth decade and then declined due to a demographic transition in the home countries;
- the degree of industrialization as measured by the share of agriculture, which indicates the (moderate) relevance of urbanisation for migration;

- the size of the immigrant stock as a measure of the “chain effects” of migration, which apparently increased over time as a determinant of migration.

However, the graphs also show that the shrinking transatlantic wage gap acted like an increasingly important brake on migration, gaining at the end enough power to bring transatlantic migration almost to a halt after nine decades. This indicates that, remarkably enough, the shrinking wage gap led close to a transatlantic equilibrium that may have stabilized itself towards the end of the period (even without the restrictions on immigration that were actually set up by the United States and other countries in the 1920s!).

What were the forces behind the closing real wage gap? Economically, three candidates come to mind:

1. migration itself as the massive transatlantic labour movements may have led to a downward pressure on real wages in the New World and an upward pressure in Europe;
2. trade in the Heckscher-Ohlin-sense, which may have led towards a transatlantic equalisation of factor scarcities and thus also to real wage equalisation, with real wages catching up in Europe and real rents catching up in the New World;
3. capital accumulation and technical progress, which may have been faster in Europe than in the New World, with a correspondingly faster growth of the marginal productivity of labour and the real wage.

Note that, of these explanations, the last one is implausible in view of what we have learnt in section 3.1.5 on transatlantic capital movements. Clearly, the flow of capital ran from the Old to the New World, and probably with it the flow of technology. Also, from FIGURES 26 and 27, we know that, from 1870 to 1913, GDP per capita as an aggregate measure of labour productivity did not grow faster in Europe than in the New World. In fact, the reverse is true so that a story in terms of a transatlantic gap in productivity growth is not attractive to account for the convergence of low European to high New World real wages.

To be sure, econometric evidence presented by the economic historians Alan Taylor and Jeffrey Williamson has convincingly shown that it is transatlantic migration that did most of the work of transatlantic real wage convergence in the period from 1870 to World War I. FIGURE 68 presents the most important of their results. The message of the table is unambiguous. In countries with a massive labour exodus such as Ireland and Italy, emigration allowed for a roughly 30 % higher real wage than would have been the case in 1910 under a counterfactual scenario of no migration. Conversely, in classical immigration countries with an originally small population like Argentina, Australia and Canada, the incoming labour re-

duced real wages by much more than 10 % by 1910; for the United States, the reduction was 8.1 %. On average, labour in the Old World gained by 8.6 % while labour in the New World lost by 12.4 %. In fact, no other explanatory variable is needed to account for the observed convergence: neither a transatlantic difference in productivity trends, which is implausible anyway, nor whatever sort of Heckscher-Ohlin-type of trade.

To be sure, the massive transatlantic labour movements of the 19th century justify to speak of an “Atlantic economy”, very much so as we speak, e.g., of a “German economy” of the 19th century, which was also characterized by a massive (internal) East-West migration. Clearly, the readiness of labour to migrate across the Atlantic at the much reduced cost of doing so gave the world economy an extent of adjustment flexibility, which was unprecedented. It is really not before the second half of the 19th century that the vast opportunities of the “frontier” in the New World were fully exploited, and this happened mostly through migration.

It is important to realize that, in this “Atlantic economy” in the period 1870-1913, the reallocation of labour in space – and maybe also the trade-integration along Heckscher-Ohlin lines – did contribute to a trend towards factor price equalization that was powerful indeed. Of course, it is extremely difficult, if not impossible, to say exactly what share of the trends we observe were due to labour migration or to trade integration or, for that matter, to capital mobility. If anything, the facts speak for the movements of labour as the decisive force. Be that as it may, the statistical picture of convergence is remarkable as the developments of real wages and of land prices show:

- International real wage dispersion – how ever measured – declined continuously and substantially as various measures in FIGURE 69 indicate, with the dispersion being defined as the coefficient of variation of real wages over a sample of 13 to 17 countries. As Kevin O’Rourke and Jeffrey Williamson have shown, it is mostly the transatlantic convergence that is responsible for this result. Taking one extreme example, we look at Irish real wages, both in absolute terms and relative to the real wages in the United States and Britain (FIGURE 70). The message is unambiguous: Ireland’s real wage – how ever defined – increased fast and continuously over the relevant period, and it did so both in absolute and in relative terms (most notably vis-à-vis the United States). Note that, in this period, Ireland did neither industrialize nor receive much foreign investment, even if the term “foreign” is economically interpreted as to include Britain. Therefore, its real wage increase is a paradigmatic case of the forces of Globalization in the form of migration.

- The price of land also converged between Europe and the New World in the period 1870-1913, thus indicating the combined forces of transatlantic migration, capital movements and trade. FIGURES 71 and 72 show real land prices for Australia and the United States (FIGURE 71) as well as for Britain, Denmark, Sweden, France and Germany (FIGURE 72). The trend for the New World is unambiguous: formerly abundant land became scarcer and dearer. For the European countries, the picture is less clear: roughly constant land prices in protectionist France, Germany and Sweden, a trend decline in free-trade Britain, and some regime switches in free-trade Denmark with a sustained rise in the early 20th century, which may reflect the high efficiency in agricultural production of this grain exporting country. All in all, however, the picture is again one of transatlantic convergence.

Finally, the factor price ratio of labour and land – the wage-rent ratio – may give us an idea of whether trade, migration and capital flows contributed towards convergence. FIGURES 73 and 74 provide some answers. For the New World countries Argentina, Australia and the United States, the trend is unambiguously downwards, which strongly supports the conjecture that abundant land became scarcer, and scarce labour became more abundant. For the Old World, the trend is upwards, but the slope of the trend is likely to be somewhat flatter. Besides, there appears to be a slight difference between the protectionist and the free-trade countries, with the free-traders experiencing the larger change.

This last result may give us a warning that it may be too early to attribute virtually all the convergence of factor prices to migration. Maybe that trade integration (and also protectionism) had a substantial influence, which is hidden in the data and statistics that are not rich enough to neatly separate the working of all the strongly intertwined forces under consideration. Hence, despite all the remarkable research done by modern economic historians on this period, Globalization in the 19th century retains some of its mystery.

3.2 The Era of Disintegration: 1914-1950

From the beginning of World War I to the end of World War II, the world economy made no progress in terms of Globalization. To the contrary, it made steps backwards: it disintegrated. The statistics show this, e. g., the most simple (though not the most precise) indicator of Globalization: the national trade share, defined as merchandise exports as a percentage of GDP (FIGURE 75). For the total of the 16 countries covered in FIGURE 75, it increased from 5.9 % in 1870 to 8.2 % in 1913, but then declined to 5.2 % in 1950, thus falling even below

the level that it had already reached 80 years before. The backlash of the interwar period was a sustained and most powerful one. From 1950 onwards, it took almost 25 years to repair it in quantitative terms: by 1973, the total trade share stood at 10.3 %, which was higher, but not very much higher than in 1913.

Note that these trends of Globalization – positive in 1870-1913, negative in 1913-50 and positive again in 1950-73 – were ubiquitous. Of all 16 countries in the table, which include all industrialized nations of today, 15 saw their national trades shares grow in 1870 to 1913 (the exception being the Netherlands), all except two saw them decline in 1913-50 (the exceptions being Canada and Sweden) and all without exception saw them rise again in 1950-73. Clearly there must have been major common forces at work that retarded the economic integration in this exceptional period 1913-50.

It is not difficult to identify these forces in actual history. In a stylized picture, the economic drama of this period can be divided into six brief time spans, which define the next sections of this manuscript: World War I and its immediate aftermath, the Great Inflation in 1922/3, a temporary stabilization with a return to the Gold Standard in the second half of the 1920s, the Great Depression 1930/2, which ushered in a period of highly protectionist policies, and eventually World War II with its own aftermath. Note that it is a drama of massive macroeconomic disturbances that provoked different economic policy responses: a brief period of what may be called cooperative liberalism following war and inflation, and a wave of un-cooperative protectionism and another war following the Great Depression.

3.2.1 World War I and its Aftermath

Right from its start, World War I set a temporary end to many institutions and configurations that had given the pre-war world economy its characteristic outlook of dynamic growth and liberal trade. The major event that symbolizes the end of “good old globalism” was the suspension of the Gold Standard by virtually all nations that participated in the war. Given the massive need for a free hand in macroeconomic war financing, gold convertibility was given up. With Britain doing this as early as 1914, a period of eleven years of floating exchange rates began and lasted until Britain’s “return to gold” in 1925.

The consequences of war financing were immediate and far-reaching. Without the external discipline of gold convertibility and with massive fiscal needs everywhere, the national money supplies expanded drastically. Technically, this happened through the channel of generous rediscounting of treasury bills and government bonds by the national central banks, which simply filled up the active side of their balance sheets with government debt and paid for their

purchases by issuing fresh banknotes. The consequence was everywhere a massive inflationary pressure, but not everywhere to the same extent. Typically, the pressure was countered by whatever means of rationing aggregate (nominal) demand, mostly some sort of administrative price and wage stops. Of course, once peacetime arrived, the pent-up liquidity was to lead either to an open price inflation or to a currency reform.

Not only on the monetary side did the massive war-time need of resources leave deep traces. In general, the share of government spending (and taxing) in GDP rose significantly. This was made possible by new forms of taxation or forced saving, which typically proved quite persistent after the war. In a sense, World War I (and to some extent also World War II) had a kind of ratchet effect: they led to a new dimension of government activity, but after the end of military conflicts, that dimension became the peacetime standard. It is not quite clear why this was so. And it is even less clear whether this is a general phenomenon or a peculiarity of a century that witnessed a powerful trend towards more public intervention in economic and social matters, with any major war working like a catalyst for a development that was to come anyway.

Note that World War I had also an important temporary impact on factor scarcity with long-standing consequences. Given the enormous need of manpower in wartime, it is not surprising that the trade union movement made major breakthroughs in many countries, notably in Britain and Germany. For the first time, unions were finally recognized as the (cartelized) voice of the workers that joined them. What had started with industrialization as an economic option – the mobilization of labour that worked and lived close together for collective action or the threat thereof (see section 3.1.1) – was now finally turned into a massive rise of power due to the widespread labour scarcity that World War I brought with it. In a country like the Germany of the 1920s, e. g., collective bargaining became the rule rather than the exception that it had been before 1914. And parallel to economic union power, the active political role of social democratic parties increased. In 1924, Britain obtained its first – though short-lived – labour-led government; and the German Social Democrats played a major role in quite a few of the governments that the unstable Weimar Republic was to have from 1919 to 1933 when Hitler came to power.

Apart from the humanitarian suffering and the tragic loss of millions of lives, World War I left a highly problematic political and economic legacy in the form of peace negotiations and treaties that followed it. At least this is how modern historians unanimously see it with the benefit of hindsight, and some contemporaneous observers such as the later famous British economist John Maynard Keynes saw it early on. In particular, the conditions of the Treaty of Versailles of 1919 and the London Agreement on reparations that followed in April 1921 were

extraordinarily harsh. For Germany, the burden of reparations was fixed at a capitalized sum of 132 Billion Gold Mark to be paid back in annual tranches of two Billion Gold Mark in cash plus various payments in kind plus a hefty tax on German exports. Note that the German GDP was 52.4 Billion Gold Mark in 1913 and 67.3 Billion Gold Mark after currency stabilization at pre-war parities in 1925. Hence the mere cash burden of reparations amounted to about twice the German GDP – an enormous sum by all standards, even if repayments were to be spread over decades.

Naturally, the London Agreement was highly controversial in the German public, but the German parliament ratified it with a narrow majority under heavy Allied pressure. The spirit of the matter can be read off the time span that was left for the ratification process: six days. Needless to say that all this poisoned the political atmosphere and played in the hands of right-wing radical parties that fiercely opposed the London Agreement. The democratic parties, who had grudgingly accepted the treaty terms, were henceforth branded as coward and timid “Erfüllungspolitiker”.

From the start, Germany did not properly meet the obligations from the treaty. The reason lay in a combination of real difficulties and lack of will. The objective problems were obvious enough. In all nations that had participated in the war, but most all in France and Germany, the economic and social challenges of demobilizing millions of soldiers were enormous. And they were further aggravated by the massive price inflation that showed up everywhere. In circumstances like these, it would have been a tough uphill struggle for any government to squeeze the economy’s absorption down by a couple of percentage points so as to finance reparations.

But sure enough, this was not the whole story. On top of it, the German authorities did not really show any enthusiasm to meet deadlines and deliver the required volumes of reparations in kind. It is therefore hardly surprising that, by mid 1922, the Germans requested a moratorium, and by early 1923, things escalated when France and Belgium, the two major reparation recipient countries, occupied the Ruhr Valley to enforce payments. This led to a passive resistance of the workforce and a virtual standstill of industrial production in the Ruhr heavy industry, which could not possibly be financed by tax receipts, although the country had just established a highly modernized Germany-wide fiscal administration. In the end, it was financed by printing money.

3.2.2 The Great Inflation 1922/23

The early 1920s were a highly inflationary period everywhere in Europe and also overseas – as a direct consequence of the liquidity that had been built up in the course of the war. However, genuine price explosions that deserve to be called “hyperinflation” – roughly defined as an inflation at a rate of more than 50 % per month (i. e. about 1.4 % per day!) – occurred in four European countries: Germany, Austria, Hungary and Poland. All four countries lived through a complete debauching of their money that was followed by a highly successful stabilisation. This unique set of almost parallel national experiences has been carefully studied by the renowned macroeconomist Thomas Sargent who, in 1983, published a seminal paper on them with the title “The Ends of Four Big Inflations”. In this section, we draw heavily on this paper’s insights. In view of what followed in the later 1920s, we focus exclusively on the German case, which was the most important one for the re-establishment of orderly international economic relations after World War I.

What happened in Germany in 1922/3? FIGURE 76 shows the development of wholesale prices in Germany from early 1919 to late 1924. The graph indicates a breathtaking destabilization, which gained momentum in the second half of 1922 and merged into hyperinflation some time in spring 1923. In the last weeks of that year, it was abruptly and completely stopped. Note in FIGURE 76 the bizarre dimension of the vertical axis, which denotes the price index for 1914=100 and runs from 1 to 1,000,000,000,000,000. In January 1919, the first point in the graph, the index stands at 262; in December 1923, it reaches 126,160,000,000,000 and stays approximately constant thereafter. In this period, the external value of the German Mark tumbled respectively as can be seen from FIGURE 77, which shows the (free market) exchange rate of the American and the German currency (in cents per Mark) from January 1920 to January 1925. It roughly tracks the path of inflation. Once convertibility was restored at the pre-war Goldmark rate of 4.20 US-Dollars, the price of the mark returned to 23.8 cents.

The mechanism underlying the inflation was very simple: the German central bank, the “Reichsbank”, discounted treasury and commercial bills, and paid with freshly printed money. Economically, the chain ran as follows: workers in passive resistance to the Ruhr occupation practically stopped working, but continued to receive their normal wage, which was paid by firms that sold commercial bills to the government, which issued treasury bills to finance its massive expenditure, and all bills were generously purchased by the central bank and paid for with new money.

In terms of the standard quantity equation $MV=PY$, with M defined as the money supply, V as the velocity of money in circulation, P as the price level and Y as national output, the process reads as follows. To begin with, M rose sharply at constant Y and V so that prices rose proportionally to M . As hyperinflation gained momentum, V also increased: once it became clear to money holders that there was an extremely fast devaluation of their holdings going on, they had every incentive to speed up spending their money so that ever more money circulated ever faster. On the other hand, the economy's output Y began to shrink because work discipline naturally collapsed: after all, there was no reason anymore to work hard when the earned money wage devalued drastically within hours so that it was more important to put all energies into smart shopping rather than diligent working. In the end, the situation became absurd: an exploding circulation of ever more money chasing a declining volume of goods. FIGURES 78 and 79 reveal the whole extent of this absurdity in 1922-3. While the increase of the money supply in 1922 and 1923 was enormous, the rise of the price level was even greater so that the money supply in real terms (in Figure 79: "real balances") declined substantially over the two years. Hence, in the middle of a flood of liquidity, money was becoming ever scarcer.

In social and political terms, the macroeconomic absurdity had an extremely dark face. It meant that those people who were not fast and clever enough to adjust to the whirlwind of hyperinflation found their real savings destroyed and their real incomes "imploding". They were the losers, and among them, there were the elderly and large parts of what may be called the honest working class. On the winning side, there were those who had early on purchased assets abroad or real estate at home and thus escaped the confiscatory inflation tax; and, of course, the quick and smart ones who were able to shop around at reckless speed while others spent their time at work. In fact, hyperinflation told a whole generation of people that honest work has no pay-off while quick opportunism can make you rich. Clearly, this left a deep-seated frustration in large parts of the population and made many people potentially receptive to radical ideas of the extreme Right and Left.

How was stabilization achieved? With the benefit of hindsight of modern macroeconomic analysis, it is clear that a complete change of the fiscal regime was needed to build up the political credibility of whatever currency reform. In a sense, one had to go back to the credibility of the Gold Standard. At least implicitly, that credibility had not only been based on the gold convertibility of the currency, but also on the promise of all governments to resort to genuine taxation (and not money printing!) to achieve a balanced budget. There is an analogy to private firms: only those firms are regarded as solvent that can credibly pretend to sell

enough of their products in the market to generate the necessary revenue to cover their costs. What the sales are for firms is taxation for the government.

It is remarkable how thorough the reforms were that all four countries with hyperinflations carried out to achieve a balanced budget. Let us again focus on the important case of Germany. The reforms of November 1923 under Gustav Stresemann, then Chancellor and later Foreign Minister, established a new currency, the so-called “Rentenmark”, which replaced the debauched old Reichsmark at the exchange rate of 1: 1,000,000,000,000 (one trillion). The “Rentenmark” owes its name to the fact that it was formally backed by real estate and land from industry and agriculture. To be sure, this “mortgage backing” had more a symbolic than a genuine economic meaning as the guarantees could hardly be expected to be redeemed by foreign investors in case of insolvencies without a massive political turmoil.

Economically, the important steps were taken in the fiscal realm. Most of all, the government budget was balanced by January 1924. On the expenditure side, public employment was cut down immediately by a hefty 25 %, and by another 10 % by the turn of the year 1923/4. Also, big public companies – notably the “Reichsbahn” and “Reichspost” (railway and mail service) – made massive lay-offs and thus drastically reduced their dependence on government subsidies. In a sense, the whole public sector was reduced to the efficient size it had before World War I, with the massive post-war sucking up of a demobilized labour force being reversed. On the revenue side, the “Reichsbank” was henceforth prohibited from any further rediscounting of treasury bills. Though this fell short of central bank independence, it did at least cut off the government from its money printing machine.

Apart from the internal fiscal and monetary stabilization, an international agreement – the so-called Dawes Plan – granted a loan of 800 Mio. Mark to Germany, thus providing an emergency cushion against potential balance of payments crises. In addition, the agreement entailed a moratorium on reparation payments, and it cleared the way to working out a new reparation scheme. This was finally agreed upon in 1930 in the Young-Plan, which drastically reduced the capitalized sum of reparations from 132 to 37 Billion Marks that were scheduled to be paid in tranches until 1988 (which, of course, never became reality after the Great Depression and its aftermath including World War II). Institutionally, the Bank for International Settlements (BIS) was founded as an organisation to survey and clear reparation payments and deal with related issues. It became the first supranational bank, located in Basle (Switzerland) where it still exists today with a much enlarged portfolio of tasks.

The German currency reform was remarkably successful. It stabilized the money supply and the price level (see FIGURE 78); it removed the odd scarcity of real balances (see FIGURE 79); and it led to a fast recovery of production (see FIGURE 80). Even more importantly, the reforms proved to be viable in the medium run: from 1924 to 1929, German GDP grew at an average annual compound rate of 4.1 %, and the price level remained stable. While the reform necessarily involved hardships, it was widely recognised as unavoidable. Most of all, the temporary rise of unemployment was accepted as an unavoidable adjustment to a sober inflation-free system of prices and wages. The German economy now stood ready to join the trend of industrial modernization that was in full swing elsewhere, notably in the United States,

3.2.3 Temporary Stabilisation

By the mid 1920s, post-war inflation had run its course everywhere, and a new period of apparently stable macroeconomic conditions was on the horizon. Therefore, it is not surprising that the return to the pre-war Gold Standard moved onto the agenda.

To be sure, the discussion about the “return to gold” had started early on after World War I – in Britain, the anchor country of the traditional Gold Standard. But in view of the strong inflationary pressures, still high unemployment and the pronounced fluctuations of the British Pound notably vis-à-vis the US-Dollar, there was still a political reluctance to impose the discipline of gold convertibility on the national economy. When the first Labour-led government of British history came to power in 1924, this reluctance, if anything, increased. A year later, however, the issue was back on the agenda as the Labour cabinet resigned and was replaced by a conservative one, the cabinet of Prime Minister Stanley Baldwin. In May 1925, Britain returned to Gold convertibility, and it did so at pre-war parity.

This decision was heavily criticized by John Maynard Keynes, at that time already a very prominent economist, who wrote a paper under the provocative title “The Economic Consequences of Mr. Churchill”. In fact, as chancellor of the exchequer (in modern terms: minister of finance), the later war-time Prime Minister Winston Churchill was politically responsible for the decision, although it was a little bit unfair to criticize him for it because he did what practically every other conservative politician would have done as well. After all, in the stabilized international macroeconomic climate of the mid 1920s, the Gold issue became more and more a question of national pride to restore British leadership than of hard-core economics – at least in the minds of Tories like Churchill (who knew very little about economics anyway).

Keynes' critical analysis came down to one central point: at the pre-war parity, the British economy was subjected to a permanent deflationary pressure that would prevent it from reaching anything close to full employment. In fact, this is exactly what happened after Britain's "return to gold". From 1913 to 1925, the price level in Britain had risen by almost 80 %, thus implying that, to reach purchasing power parity in terms of gold, an 80 %-increase of the gold price (and an appropriate devaluation of the British Pound) would have been appropriate. Things were even made worse by the terms at which other large countries were to join the Gold Standard in 1926: the US at the pre-war parity, but after a much lower post-war inflation than the one in Britain; Germany at the pre-war parity, but after a currency reform that had "purged" the new German Mark of all post-war (hyper-)inflation; and France at 20 % of the pre-war parity, which by far overcompensated the internal price effects of inflation notably vis-à-vis Britain. Hence, in real terms, the British Pound was vastly overvalued vis-à-vis the currencies of all major trading partners. The reverse was true for the French Franc and possibly the US-Dollar.

Industrial unemployment rates in the 1920s support this conjecture (see FIGURE 81): in the second half of the 1920s, about 10-12 % of all industrial workers in Britain were unemployed, compared to usually less than 5 % in France and the United States and 7-9 % in Germany. Note that the Scandinavian countries Denmark, Norway and Sweden had also very high unemployment rates. Just as Britain, they joined the Gold Standard at pre-war parities and experienced a massive overvaluation of their currencies. Note also that the reverse holds for Belgium and the Netherlands, which followed the "French model" of undervaluation.

From all this, it becomes clear that the return to the Gold Standard had two quite different faces. On the one (and positive) side, it provided again a framework of liberal international finance that had a long history of success and a great degree of credibility attached to it. After the years of macroeconomic turmoil or even chaos in some places, it signalled the return to a degree of order and stability that opened up capital markets anew. Notably France and Germany, the two countries battered most by wartime losses and post-wartime disturbances, profited from it. In the case of France, a massive current account surplus with massive inflows of gold at high degrees of employment and capacity utilization made for a comfortable macroeconomic constellation. In the case of Germany, a current account deficit and massive infrastructure and modernization investments could be financed on international markets, mostly by American capital, thus making the burden of reparation duties tolerable. In fact, more than that: in the period 1924-9, the inflow of capital into Germany was roughly twice as high as the outflow in terms of reparations.

On the negative side, however, there were Britain's persistent macroeconomic problems. The country that had quite easily stood firm as the anchor of the pre-war Gold Standard was grunting and groaning under the post-war Standard with its new (real) exchange rates. Apparently, Britain was not anymore the leading economy that generated large current account surpluses and could thus serve as the world's main provider of capital and lender of last resort. The only country that could have been able to take up this role were the United States. But on the other side of the Atlantic, there was not yet any political will to do so. As a consequence, some potentially dangerous asymmetries piled up, which might have been bearable in normal times. But economic history soon showed that times were anything but normal.

3.2.4 The Great Depression 1930/32

What happened in the world economy in the years 1930 to 1932 deserves the name "Great Depression". By all standards, it was the worst worldwide breakdown of economic activity in peacetimes up to that date and, remarkably, up to the present. FIGURES 82 gives a quantitative impression of the depth of the fall, in terms of national output (GDP) and industrial production for 21 countries. From 1929 to 1932, output shrank massively in the vast majority of countries, with the worst contractions taking place in the United States, Germany and Austria. The slump hit all sectors of the national economies, but it was worst of all in industry. In all countries in the table, the decline in industry turned out sharper than in the rest of the economy, and in the United States, Germany and Austria (and also Poland), it reached altogether disastrous dimensions. After all, a drop of industrial production by 45 % as in the United States or 41 % in Germany practically means a standstill of all those industrial machine capacities that are typically used to produce goods beyond the mere subsistence items of daily consumption like food and clothing.

The disaster left everywhere deep traces in the labour markets. Recall FIGURE 81, which provides industrial unemployment rates for eleven countries. Although the data must be interpreted with caution due to deficiencies in comparability and coverage, it does convey a quite uniform message: from 1929 to 1932, industrial unemployment rose massively everywhere, with the worst turn to the bad reached in those countries that experienced the most disastrous production decline. The peak rates of industrial unemployment were reached in 1932-Germany with 43.8 %, after a low of 8.6 % in 1928, and 37.6 % in the United States of 1933, after a low of 5.3 % in 1929. For the economy as a whole, it is not unrealistic to assume that, at the depression trough, at least one out of four Germans and Americans was out of work.

Among economists and economic historians, there is unanimous agreement that the downturn in 1929/30 had its origin in the American economy and was transmitted to Europe through a variety of different channels, which are closely linked to the integration of international capital markets. It is therefore appropriate to tell the story of the Great Depression by first looking at the United States and then moving on to the rest of the world, with a particular focus on hard-hit Germany, which was the country whose economy suffered first and foremost under the American path of events.

For the economy of the United States, the 1920s – and notably the second half of the decade up to 1929 – were a period of excellent macroeconomic performance. Industrial production grew at an average annual compound rate of 5.2 %, and the economy-wide unemployment rate – when measured with the methodology used nowadays – was about 5 %, i. e. at a level that has been characteristic for the American economy in boom times ever since. Remarkably enough, it was not a time of strong inflationary pressures. In the second half of the decade, consumer and producer prices were stable or, if anything, declined slightly. Hence, in the markets for goods and services, there was no indication of overheating. Apparently, the supply side of the economy was elastic enough to accommodate powerful waves of consumption and investment demand that came with the first big tide of motorization in the world.

Note that it was the time when the Ford Motor Company launched the famous Model T, which was to become the first car in the world for the broad middle classes (and not, as previous cars had been, an item of snobbish consumption pleasures for a tiny rich minority). With its industrial backward and forward linkages, motor car production was a most powerful engine of economic expansion. And so were the new production methods that came with it. E. g., Henry Ford's assembly line, which established a new industrial philosophy of exchangeable parts, was soon imitated in other branches of manufacturing and became a major force that fuelled American productivity growth. Sure enough, the United States became a much admired industrial powerhouse, and the new fashionable address for many visitors from Europe who wanted to know more about upcoming products, processes and techniques. Clearly, what had been Britain for a long time in the 19th century now were the United States: the shining example of modern industrial technology.

Against the background of this powerful dynamics, however, the industrial expansion had at least one critical side: while goods prices remained stable, asset prices boomed, both in the market for shares and for real estate. FIGURE 83 presents some numbers on this boom: between 1926 and 1929, the so-called Standard Statistics Index of the New York Stock Ex-

change more than doubled, with almost half of the share price rise taking place from summer 1928 to autumn 1929. Whether this increase should be considered as a “bubble” in the sense of a speculative revaluation that goes well beyond business fundamentals remains an open question, even with the benefit of hindsight about 80 years after the events. Be that as it may, the increase was sharp and strong, though not altogether unique in the history of share pricing.

What was exceptional, however, was the reaction of the American central bank, the Federal Reserve System (henceforth briefly: the “Fed”). In modern capital market terminology, we would say that the Fed “leaned against the wind” of asset price inflation. It raised discount rates quite substantially to counter the speculative movements of the market: from 1925 to 1929, the rate was put up in successive steps from 3.5 % to 6.0 % in 1929, a level that was a peak since 1919/20 and remained a peak until 1969/70 (!), when massive inflationary pressures and a weak Dollar were responsible for another sharp tightening of monetary conditions. Note that, at stable goods prices in 1929, a 6 %-discount rate translated into a much higher real rate of interest than a decade earlier or four decades later. Hence, by historical standards, the policy of the Fed in the late 1920s must be regarded as unusually tough and restrictive.

How did money markets react? FIGURE 84 shows monthly statistics for the period 1919-39 of five different short-term interest rates in the United States. The overall message is unambiguous: all rates peaked in 1929 more or less parallel to the peak in the discount rate. Hence it is reasonable to conclude that the leeway for speculative engagements was severely narrowed. Apparently, a situation of extreme tightness was reached, and that unloaded itself on so-called Black Friday of October 29, 1929, in a dramatic stock market crash. Not until so-called Black Monday of October 19, 1987, did the New York Stock Exchange experience again a downward correction of share prices to a comparable magnitude in one trading day.

Among economists and economic historians, it is today uncontroversial that the crash in share prices did not by itself “cause” the Great Depression. However, it did mark a substantial change of macroeconomic parameters. Most of all, it changed public perception and monetary policy: the discount rate was lowered immediately, and further cuts followed down to 2.5 % within one year. Visibly to everybody, the boom was over though it remained not clear at all for months how severe the downturn would be. By early 1930, a brief recovery of share prices made things look not at all hopeless. However, the further economic news during the year 1930 made it increasingly clear that something unusually bad was going on.

House prices fell and building activity contracted sharply. Agricultural prices tumbled and farmers faced bad business prospects – in a country where about 20 % of the labour force was still working in agriculture. Most importantly, automobile production slumped: from its peak of 4.5 million cars in 1929, it fell to 2.8 million in 1930, 1.9 million in 1931 and 1.1 million in 1932, an overall cut by 75 % in three years. Note that it was not before 1949 that American car production reached again its first peak of 1929.

In the last quarter of 1930, the downturn reached a new dimension as the rate of bank failures rose dramatically. From this time on, it was clear to everybody that the economy was entering a most severe recession, if not a genuine depression. In early 1931, some indicators seemed to point towards a slight recovery, but things worsened again as further bank failures occurred. In addition, the crisis began to have an ever stronger international dimension, with a negative feedback on the American economy. In particular, Germany pursued an extremely tight fiscal and monetary policy and restricted international trade and capital movements by introducing exchange controls. And, in September 1931, Britain left the Gold Standard, which relieved the country from balance of payments pressures, but shifted the burden of external restrictions on other countries, notably the United States, where – in the middle of a vigorous downturn – short-term interest rates went up sharply (see FIGURE 84 for the later year 1931).

Eventually, the cumulative decline petered out. This happened some time in 1932 when the depression reached its trough and gave way to a sluggish recovery that became visible in the course of the year 1933. Politically, the presidential elections of November 1932 led to a replacement of the unfortunate Herbert Hoover by Franklin D. Roosevelt who engineered the policy of the so-called New Deal. It entailed, from 1934 on, a more expansionary policy mix with a significant increase of public spending and a rise of the money supply, combined with a devaluation of the US-Dollar vis-à-vis gold. With this new policy stance, the worst part of the depression was over, although it took the American economy another three years until national output (in real terms) reached again the level of 1929, and it did so at a much higher rate of unemployment than before the crisis.

FIGURE 85 shows the quantitative profile of the Great Depression for the United States in terms of annual data of some major macroeconomic variables. On the real side of the economy, the picture is bleak. From 1929 to 1932/3, the unemployment rate rose from about 3 % to 25 %, GNP (gross national product) declined in real terms by about 30 %, consumption by roughly 20 %, and investment virtually vanished. By 1932/3, hardly any firm in America was ready to expand its capital stock as the degree of capacity utilisation was so bad and the

expected rate of return so low that there was no point in planning new production halls or installing new machinery. Bluntly speaking, the economy had been reduced to what may be called its “consumption core”. Note that, despite this fact, government spending in real terms remained roughly constant over the period 1929-32/3. It did not rise before Roosevelt’s New Deal set in, with government purchases increasing sharply from 1934 on. During the depression, however, there was no conscious effort made to counteract the economic contraction in the private sector by an expansionary fiscal policy.

On the monetary side of the economy, the picture is equally remarkable (see again FIGURE 85). In the period 1929-32/3, nominal interest rates declined, though with a slight, but economically important upward kink in 1931/2, which was due to the medium-term effect of Britain’s leaving the Gold Standard. The (nominal) money supply contracted substantially, by roughly a quarter between 1929 and 1933, with a particularly strong decline of 12.4 % in 1931/2, probably also a consequence of the external pressures put on the Fed by Britain’s turn away from gold. Parallel to the contraction of the money supply, the price level declined by more than 20 %, and that happened mostly in the years 1931 and 1932 with annual deflation rates of about 10 %. Note that, due to the roughly proportional decline of the nominal money supply and the price level, the money supply in real terms (in FIGURE 85: “real money balances”) remained more or less unchanged all throughout. Not before the end of price deflation in 1934 did it rise again.

So much for the macroeconomic profile of America’s Great Depression. How can we explain this disaster? Among economists and economic historians, two major types of explanations have been advanced. One of them is called “spending hypothesis”, and it goes back to a number of early Keynesian economists (including John Maynard Keynes himself); it was rigorously developed by the economic historian Peter Temin in a seminal book published in 1975. The other one is called “money hypothesis”, and it was stipulated by Nobel Laureate Milton Friedman and Anna Schwartz in their famous book “A Monetary History of the United States” published in 1960. In a highly stylized form, both hypotheses can be portrayed in a standard IS/LM-framework that describes the interaction of asset and goods markets in an economy with fixed prices. Note that, as we have seen from FIGURE 85, the assumption of fixed prices is descriptively false, but it serves as a good analytical starting-point. We shall drop it later on.

FIGURE 86 presents the two hypotheses in a graphical illustration of an IS/LM-model in output/interest rate-space with the (downward sloping) IS-curve as the geometric locus of equilibria in the goods market (or more precisely: equality of saving and investment), and the LM-

curve as the geometric locus of equilibria in the money market (in a world with one liquid asset: money). The spending-hypothesis basically says that the Great Depression can be described as a massive shift of the IS-curve to the left (FIGURE 86a), which was due to a drastic reduction of private spending in the wake of the stock exchange crash in autumn 1929. In turn, the money hypothesis holds that the Great Depression can be described as a massive shift of the LM-curve to the left (FIGURE 86b), which was due to the drastic contraction of the money supply in the wake of bank failures in autumn 1930.

Given the stylized facts sketched above, both explanations have their strong and weak points. In the case of the spending hypothesis, it is fully consistent with the simultaneous decline of output and the interest rate that can be observed over most periods of time during the depression. On the other hand, it places an enormous weight on some sort of “autonomous” spending shock to start from, and that makes only sense if one assumes a massive negative wealth effect on consumption and investment that was brought about by the stock market crash and the related downward correction of asset values in all balance sheets of the economy. Even in view of the prior asset market boom, however, this looks a bit far-fetched. To be sure, it may be enough to explain a business downturn or a standard recession, but certainly not a Great Depression.

In turn, the money hypothesis has the potential to explain what happened in the course of the banking crises in 1930 and 1931 and with Britain’s leaving the Gold Standard in September 1931. In other words: it may possibly account for the depth of the crisis in its later stages – as a consequence of a massive contraction of the money supply. Unfortunately, however, it is flatly contradicted by the crucial fact that the interest rate did not rise when output fell in the wake of the banking crises, thus making it hard to believe in a leftward shift of the LM-curve. To be sure, short-term interest rates increased only in September 1931 when Britain went over to free floating, which is consistent with the money hypothesis; however, interest rates did not increase – in fact, they continued to fall – in the course of the American banking crises from autumn 1930 to spring 1931 (see FIGURE 84 with October 1930 marked by a vertical line).

Given these bright and dark spots of both hypotheses, it is tempting to search for an explanation that combines the power of both, but avoids their pitfalls. Such an explanation may be called an “augmented” money hypothesis, and it focuses on the crucial observation that, from 1930 onwards, the price level declined and a powerful deflation set in. This means that the simple IS/LM-framework, which assumes constant prices, reaches its limit, notably with respect to explaining investment demand. As a major – and empirically the most volatile –

component of aggregate demand, investment depends on the expected real interest rate (r). And this comes down to the nominal interest rate (i) only if investors expect price stability. Let π be the expected rate of price inflation, then the expected real interest rate r is given by $r = i - \pi$. And for $\pi < 0$, as most probably was the case once price deflation was firmly established in 1930-32, then the real interest shot up despite a very low nominal rate i . In fact, the annual rates of actual deflation in 1932/3 of roughly 10 % p. a. suggest that expected real interest rates were becoming very high indeed, maybe in the range of 11-13 % p. a. Hence clearly, investment demand could be affected, and thus a massive leftward shift of the IS-curve induced, which can be graphically described just as in FIGURE 86 (a).

What makes this explanation an “augmented money hypothesis” – despite the shift of the IS- and not the LM-curve – is the deep cause of the shift. After all, it is induced by the dynamics of deflation, and that was the outcome of a drastic fall in the nominal money supply that was not counteracted by an appropriate action of the Fed. Banking crises happened and led to a contraction of money circulation, and the Fed allowed this contraction to swell up to a genuine deflation. In the end, it was this deflation which kept the real money balances roughly constant, but this happened at the enormous cost of a virtual vanishing of investment and an awful shrinkage of output. Here, clearly, lies a major failure of monetary policy, which contains many lessons for the future that followed the Great Depression.

As in our account of German hyperinflation in 1922/3, an interpretation in terms of the quantity equation $MV=PY$ is helpful. Recall that, in this equation, M denotes the money supply, V the velocity of circulation, P the price level and Y the output level of the economy. In the United States in 1930/2, the initial downturn was caused by an autonomous contraction of spending, meaning that V and Y declined and M and P remained constant. Beginning with the banking crises in the last quarter of the 1930s, M declined, and so did, as a consequence, P and then Y , thus further deepening the depression through monetary channels that fed back into spending cuts. In the end, output stabilized at a very low level, which we have called the “consumption core” of the economy.

All in all, a major political responsibility for the disaster must be allocated to the central bank and the government of the United States. Neither of them saw the urgent need to counteract the deflationary spiral by increasing the money supply and/or increasing public spending. Why this was so must remain open. In view of the public discussion at that time, however, it is safe to say that the very idea of a “Keynesian”-type stabilisation policy was neither analytically developed nor politically realistic. With the benefit of hindsight, it is clear that bad mistakes were made. But if Keynes’ famous contention is true that the world is ruled by ideas of

economists and political philosophers – and little else, then it is hard to put too much blame on the main actors in the American drama of the Great Depression.

The breakthrough to new ideas on short-run macroeconomic stabilization was made in Keynes' "General Theory of Employment, Interest, and Money", which was published a few years later in 1936, exactly against the background of the most recent disastrous experience of depression and mass unemployment. Apparently, it needed a major macroeconomic failure to give this seminal book the necessary public attention to make short-run monetary and fiscal stabilization a kind of staple product in the toolbox of policy makers. As we shall see in section 3.3 of this manuscript, this staple product was then misused in circumstances that were not at all deflationary. Thus, in the 1970s and early 1980s, price inflation was fuelled, and with it came a new political and academic interest in the limits of Keynesian macroeconomic "fine tuning".

So much for the Great Depression in the United States. Let us now turn to the rest of the world. Of all nations, the most directly affected by the events across the Atlantic was Germany, the country that depended strongly on American financial capital as described in section 3.2.3. In the second half of the 1920s, a kind of post-war "triangle" of financial relations had emerged with three main players: Germany, France and the United States. It worked as follows: Germany paid reparations to France, France repaid wartime debt obligations to the United States, and the United States granted loans to Germany. This triangle worked as long as American capital continued to flow on a medium and long-term basis to Germany. It was precisely this that began to change when the speculative boom gained momentum in the United States in the late 1920s.

What happened exactly? With American money markets tightening in 1928/9, the refinancing of German loans became ever more short-term, thus exposing borrowers in Germany to increasingly high risks. With the stock market crash, these risks materialized as transatlantic capital was called back in an attempt to improve the liquidity position of American lenders. This led to a massive and persistent burden on the German balance of payments and, within the confines of the Gold Standard, to a highly restrictive stance in monetary and fiscal policy. In the prevailing circumstances, however, all deflationary policy attempts were doomed to fail. Banking crises in Germany and in neighbouring Austria in 1931 were the reason for a major international attempt at relieving the burden for Germany, culminating in the so-called Hoover-moratorium, which stipulated a one year break of German reparation payments from mid 1931 to mid 1932 but actually terminated them altogether as we know with the benefit of hindsight. In addition, a loan of 420 Mio. Mark was granted to Germany by an international

consortium that consisted of the United Kingdom, France the United States and the newly founded Bank for International Settlements (BIS). As a matter of fact, the measures proved insufficient to prevent further massive banking failures in July 1931. As a consequence, Germany restricted the convertibility of its currency by administrative means, thus indirectly suspending the Gold Standard, but without any general stimulating effect of a genuine currency devaluation. On top of this, the deflationary policy was further tightened by the government of Chancellor Heinrich Brüning, with compulsory wage and price cuts that were imposed by decrees, which used all executive emergency options of the German constitution to the limit.

In the end, the German government had carried out a massive deflationary programme, backed up by bureaucratic external exchange controls. The results read like this: between 1929-32, wages declined by 25 %, consumer prices by 22 %, producer prices by 30 %, industrial production by 40 % and GDP by about 16 %. The unemployment rate rose from 9 % to over 30 %. It is no surprise that this unique set of deliberate deflationary policies and their results earned Heinrich Brüning the sarcastic nickname "Hungerkanzler" (Chancellor of hunger). In any case, the disastrous state of the German economy in 1932 paved the way for the rise of Hitler to power though, tragically enough, the first indicators of the business cycle pointed upwards a few weeks before the Nazis took over.

Note that, as in the American case, it is very hard to blame the acting politicians at the time for the disaster. They did probably no more than what most others would have done in view of the economic orthodoxy that prevailed at the time. In the German case, there is even a major historical antecedent that may help to explain the policy stance in 1930/2: the Great Inflation in 1922/3. After all, a whole generation of Germans had witnessed their savings and wealth being destroyed in the course of a hyperinflation. In these circumstances, it is not quite so surprising that a government and a central bank regarded their hands tied to fiscal and monetary orthodoxy. Even with the benefit of hindsight, it is hard to say how the capital markets would have reacted if the public authorities in Germany had deliberately opted for an expansionary policy outside the realm of the Gold Standard, which gave the country's macroeconomic stance that minimum amount of credibility that seemed to be urgently needed in international transactions.

In this respect, other countries – and notably the United Kingdom – faced a less dramatic policy dilemma. In the British political debate, leaving the Gold Standard could be seen as a rather pragmatic question of sacrificing some national pride as the traditional hub of world finance for a considerably extended leeway for expansionary policies. In view of dwindling

gold and currency reserves and persistent deflationary pressures, Britain eventually opted against the external constraints and in favour of internal stabilisation. In September 1931, Britain left the Gold Standard – a step, which was widely criticised in the British and the international public, but fully supported by some critical observers, notably the economist John Maynard Keynes. In the wake of the United Kingdom, all Scandinavian countries and all British dominions opted out as well. As early as 1929/30, this step had already been done by quite a few Latin American and overseas countries that were hard hit by the world-wide slump of agricultural prices. Thus, after 1931, only the United States and the so-called Gold block in Europe (France, Belgium, the Netherlands and Italy) remained within the Gold Standard. All of them devalued their currency later-on, beginning with the United States in early 1933 and Italy in 1934 and ending with France and the Netherlands in 1936.

FIGURE 87 gives a summary picture of the “devaluation race” that took place in Europe in the 1930s. The dates of “gold exit” or devaluation make clear that, by the second half of the 1930s, the Gold Standard was dead. Looking over its stepwise decay, questions arise. Was that decay a good or a bad thing? Would it have been better to stick to gold and fixed parities? Or were there other superior policy alternatives? In a paper with the suggestive title “Exchange Rates and Economic Recovery in the 1930s”, Barry Eichengreen, an economic historian, and Jeffrey Sachs, an economist, have given some theoretically based empirical answers to these question. In the following paragraphs, we draw heavily on their line of reasoning.

For a start, one has to distinguish two different effects of a currency devaluation, i. e. by lowering the gold price of its currency in the Gold Standard:

- By devaluation, a country redirects international demand for goods and services towards its own supply of tradable goods. In times of depression, this is aptly called a beggar-thy-neighbour-policy, i. e. a zero-sum-game in which nobody can win anything other than what he obtains in redirected demand from others. Demand is diverted and not created.
- By devaluation, a country re-values its own gold and foreign exchange reserves. This gives it the leeway for a more expansionary monetary policy, thus widening the external constraint. This happens at nobody’s expense. In times of depression, it has the character of a positive-sum-game because others will not suffer and may even profit from the unilateral demand increase. Demand is created and not diverted.

Hence a unilateral devaluation is a combination of demand diversion and demand creation. From this insight, it is easy to arrive at a first best solution. A cooperative devaluation of all

currencies within the Gold Standard by the same rate vis-à-vis gold would give every country in the system more leeway for an expansionary monetary policy, but leave relative price levels of traded goods between the countries unchanged. Thus the demand diversion would be avoided, but the gains from monetary expansion in terms of demand creation could be reaped. Thus, at least theoretically, there was an easy exit out of the external dilemma of the Gold Standard under the conditions of depression in the 1930s: get together at an international conference (an “economic summit”) and go for an adequate and coordinated rise of the price of gold in terms of all national currencies. Of course, this theoretical solution requires that the temptations of a classical prisoner’s dilemma be overcome by cooperative politics. It is doubtful whether this was feasible in the uncooperative international spirit of the early 1930s.

However, if a cooperative effort was not politically feasible, it is far from clear – as many contemporaneous critics suspected – that the uncoordinated devaluation race as it actually took place was worse than sticking to a depression-prone system altogether. From a worldwide perspective, it certainly was better because the net balance of demand creation and demand diversion is positive since the latter cancels out between countries. From a single national perspective, however, things may look differently – at least for the remaining countries in the Gold Block who refrained from devaluation for a very long time. Be that as it may, it would be misplaced to discard what happened as a solely negative example of beggar-thy-neighbour-policy. In the prevailing circumstances, it may have had the character of a badly needed emergency exit out of a system that pulled everybody down. And, in the end, it may have shifted the pressure onto the shoulders of those who still stayed within the system, and thus force them to leave it as well. After all, this may have been the only politically feasible way to obtain the necessary external flexibility to draw Europe and America out of the depression.

How successful were the countries that left the Gold Standard in terms of their subsequent macroeconomic performance? FIGURES 88 and 89 give a rough-and ready answer. They show for the year 1935 the level of industrial production (FIGURE 88) and the export volume (FIGURE 89) for ten European countries plotted against the exchange rate, and all variables normalized on the pre-depression level (1929=100). The message of FIGURE 88 is unambiguous: those countries that devalued drastically (and early), notably Britain and the Scandinavian countries, achieved a significant rise of industrial production. In turn, those of the Gold Block did not. A similar link is to be identified in FIGURE 89 between exchange rate and export volume. Clearly, devaluation was a quite successful instrument to pull a country out of the deep recession. Or, to put it differently: the chains of the Gold Standard can be held responsible for at least part of the depth and length of the Great Depression in the 1930s.

3.2.5 Paths towards Autarky

The 1930s were a time of rising protectionism. It all started with the slump in the world markets for agricultural products. FIGURE 90 shows that an index of raw material prices in world markets declined substantially from 1925 to 1928/9 only to slip into an even steeper fall thereafter. The trough was reached in the middle of the Great Depression in 1932, with the index standing at less than one third of its level of 1925. Note in FIGURE 90 that the price slump hit virtually all markets, be it cotton or silk, rubber or tin, tea, sugar or wheat. The reason was at first a rapidly increasing supply, but later, with the depression holding sway, a drastic fall of demand.

With the slump in prices, the main exporters of agricultural products ran into massive balance of payments problems. This is why, in 1929/30, most of the Latin American countries as well as Australia and New Zealand left the Gold Standard or devalued their currencies drastically. Besides this, there was the latent danger of a new tide of protectionism, which could at first be held at bay in a world economic conference in 1927 and another one in 1930. At that time, however, the United States were already preparing a massively protectionist bill that passed Congress in June 1930 as the Smoot-Hawley-Act. In essence, it consisted of a sharp rise of tariffs for agricultural produce to protect America's own farmers who were also badly hit by the world price slump.

FIGURE 91 shows the average US-tariff rate from 1915 to 2000: with the passage of the Smoot-Hawley-Act, the level of tariff protection rose to almost 60 %, an all time high by American standards. Other countries followed suit immediately, and a retaliatory trade war set in. Within a short period of time, Europe, America and the rest of the world became fortresses of protectionism. As a consequence, world trade "imploded". FIGURE 92 shows the monthly volume of world trade for the period from January 1929 to March 1933. As the "spider web" nicely shows, the trade volume shrank month by month. Starting from a high of almost 3 billion \$ in January 1929, it ended at a low of less than 1 billion \$ in March 1933. To some extent, this disastrous contraction may be due to the general business downturn, but a major part of it must surely be booked on the account of protectionism. After all, trade did not recover to its prior level even after the Great Depression was over.

In fact, things became even worse in some countries, notably in those where right-wing dictatorships were established. They imposed deliberate regimes of autarky. E. g., the National Socialists in Germany pursued a policy of central control of international trade, combined

with strong efforts to replace imported raw materials by domestically produced substitutes that were to be developed by the German chemical industry. The spirit of free trade and mutually advantageous exchange had gone. It was not to return before the end of another world war.

3.2.6 World War II and Its Aftermath

Economically, World War II marks the climax of disintegration of the world economy. With the benefit of hindsight, however, it also marks a big turning-point to the better. While the number of casualties of military conflicts and horrible crimes against humanity (notably the holocaust) reached a frightening peak in the course of the later stages of World War II, it was exactly at this time that the first diplomatic efforts were made to design a new post-war political and economic order. And, at least in the western hemisphere, this new order was to mean a more open economy and society that stood in the tradition of the successful pre-World War I period.

During World War II, market mechanisms of resource allocation were largely replaced by planning and command by public authorities, as usual in war times. This happened in most countries involved in the huge military conflict. And in these places, it was accompanied by an inflationary financing of the war (often combined with price controls) that allowed a maximum short-run mobilisation of resources but was clearly unsustainable in the longer run. This is why, after the end of the military conflict in 1945, most countries went through some sort of “demobilisation crisis”, which lasted for a few months up to three years, depending on the particular circumstances.

Once again, the case of Germany was particularly extreme and important, just as it had been after World War I, but with a completely different long-term outcome. After the unconditional surrender in May 1945, the country was occupied and its economy strictly controlled for a while. In June 1948, a currency reform was carried out in those parts of the country that were under American and British command. The reform was successfully combined with a balancing of public budgets, a strict control of the money supply by a newly founded independent central bank and, most importantly, the liberalisation of prices. This reform turned out hugely successful. It led to a fast reallocation of resources away from war-time demands towards peacetime production lines of consumption and investment goods, which were urgently needed. And all this happened with market prices regaining their proper economic role of indicating relative scarcities, but the price level remaining remarkably stable. In economic and psychological terms, the reform prepared the ground for the so-called German economic

miracle of the 1950s, which stood in stark positive contrast to the country's unhappy record of the post-war period.

A similar, though not quite as spectacular development took place in other European countries. We will briefly describe and analyze this common rapid post-war growth of Europe in the following section 3.3. Note, however, that it was largely confined to those countries that were not under Soviet domination. In fact, to the east of what became to be called the Iron Curtain, a system of tight state control and planning was established, which prevented that part of the continent from integrating into world markets. Not before the fall of the Iron Curtain in 1989/90 did the eastern half of Europe, which included East Germany, have the opportunity to join the train of Globalization that was then entering a new stage. We will return to this matter in part II of the lecture.

3.3 The Second Era of Integration

3.3.1 Rapid Post-War Growth 1950-73

From the early 1950s to the early 1970s, economic growth in Europe and in the world as a whole was exceptionally fast. FIGURE 93 provides some numbers to put this growth into a comparative historical perspective. Real GDP of the twelve European countries covered in the table grew by 4.6 % p. a. in 1950-73. This is by far the highest growth rate in all comparative periods from 1890 to 1992, easily distancing even the respectable 2.6 % of 1890-1913. In terms of productivity growth, the performance is even more spectacular: real GDP per capita and per person-hour grew at least two percentage points faster (p. a.) than at any comparable period since 1890. A side view to FIGURES 26 and 27 shows that this exceptional performance – although most striking in Europe – was not confined to the old continent. In particular, Latin America, Africa and eastern Asia grew fast as well, with Japan reaching altogether staggering real GDP growth rates of 8-10 % p. a., depending on whether we measure in absolute or per capita terms.

What were the forces behind this rapid post-war growth? At least with respect to war-damaged Europe, it may be tempting to think of the growth performance being something like an extended reconstruction that lasted for up to two decades. However, any such hypothesis is easily falsified by a more careful examination of the time when most European countries reached again the highest level of GDP that they had ever attained before World War II (see FIGURE 94). Column (2) of the table clearly indicates that all countries had reached that level by 1951 at the latest. Even for the major “war losers” Austria, Germany

and Italy, the time of mere reconstruction did not last any longer, although they had been thrown back through war destruction and subsequent occupation to GDP-levels of 1886, 1908 and 1909 respectively (see column (1)). Apparently, the period of mere repairing of war-time damages was a most successful one as the very high growth rates from 1945 to the pre-war-break-even-year in column (3) in FIGURE 94 show. But being a most successful period, it was also a very short one so that all the rest of growth up to 1973 must be seen in a different light.

While the idea of mere reconstruction as the driving force of growth is not convincing, a notion of “growing into one’s potential” may come much closer to the truth. FIGURE 95 shows why. It presents a per-capita growth ranking of 16 European countries for the period 1950-73, and that ranking delivers the message that two types of countries grew fastest:

- those central European countries that had performed badly in the interwar period, but were already well on the way to the top income league before World War I, notably Germany, Austria and maybe Italy (with its growth performance being concentrated in the north of the country);
- and those countries notably at the periphery of Europe, which had been left behind anyway and went through a belated process of industrialization, such as Norway, Ireland and Portugal.

On the other side were countries like the Netherlands and the United Kingdom as well as Switzerland and Sweden, which had reached high income levels early on and/or profited during the wars and the whole interwar period from a status of political neutrality. Apparently, some “growing into one’s potential” was going on in Europe at that time.

In terms of full employment of the labour force, the period 1950-73 was also an exceptionally successful one. FIGURE 96 provides some numbers for a sample of European countries. Comparing important benchmark years (1929, 1938, 1950, 1973 and 1989), the year 1973 stands out as that point in time with the lowest average rate of unemployment of 2.9 % for twelve European countries. Looking over extended periods in between (and before or after) the benchmark years, the bright picture is fully confirmed: an average unemployment rate of 2.5 % for four large European countries (France, Germany, Italy and the United Kingdom), much lower than any level that had been reached since World War I and was to be reached up to the present.

Note that the period 1950-73 was also one of fast trade integration. Recall FIGURE 75, which showed trade shares for 16 countries in the western hemisphere. The shares all grew from 1950 to 1973, with the total increasing from 5.2 % in 1950 to 10.3 % in 1973. Hence, by

the early 1970s, the degree of Globalization – if simply measured by the trade share – did eventually surpass the pre-World-War I level of 8.2 % in 1913. It is therefore fair to say that the traces of disintegration that the interwar period had brought about were finally removed by the early 1970s. Remember, however, that this is true for the western world and notably western Europe and North America, but certainly not for the centrally planned economies of eastern Europe, which remained largely isolated from the rest of the world.

3.3.1.1 The American Lead

In the late 1940s and early 1950s, the United States were the absolutely leading economic power in the world. GDP per capita – measured in purchasing power parity terms – was about twice the total western European level (see FIGURE 29). From this statistical fact alone one can see what two world wars plus an economic disintegration in the interwar period had left as legacy: a huge transatlantic gap. Not only had the economic leadership been shifted from the United Kingdom to the United States, but the fast catching up of many continental European regions had been stopped and partly reversed for a whole generation.

This change was dramatic, not only in statistical and economic, but also in political terms. It meant that it was up to the United States to take a major responsibility to lead the world economy back onto a path of integration. This was exactly the task that Britain was willing, but not anymore able to meet in the interwar period, with the United States still taking an isolationist stance that found its most uncooperative expression in the passing of the highly protectionist Smoot-Hawley Act in the early 1930s. In the late 1940s, however, the political climate in the United States had definitely changed towards a more positive view of American international responsibilities. While there were still occasional backlashes into unilateral isolationism, the beginning worldwide conflict of ideologies between the free and democratic West on the one side and the communist East dominated by the Soviet Union on the other made the American public ready to accept an active engagement. It was this new constructive spirit that was of utmost importance to set the points for a powerful economic resurgence of western Europe in an increasingly integrating world economy.

3.3.1.2 Others' Catching Up

The statistical picture of Europe's catching up vis-à-vis the United States is altogether impressive in the period 1950-73. FIGURE 97 shows econometric estimates of trend growth rates of output per person for seven European countries and the United States in four historical time spans including the one under consideration. In the period 1951-73, labour produc-

tivity grew by an annual average ranging from 3.4 % in Sweden up to 5.3 % in Italy, as compared to a 1.5 % p. a. increase in the United States. With 2.2 % p. a., only slow-growing Britain fell somewhat behind the European average, but still surpassed the productivity growth performance of the United States. Hence, there was a powerful and ubiquitous European catching-up going on, which stood in sharp contrast to the much more diversified statistical patterns of earlier and later periods (see also FIGURE 97).

To be sure, the catching-up was not confined to a mere growth of labour productivity and per capita output. FIGURE 98 gives a sample of different variables that are important for the questions and the period at hand. The message is unambiguous: in terms of national income, industrial production and the volume of foreign trade, most European countries grew much faster than the United States. And within Europe, those that caught up most quickly were those that had most underperformed in the interwar-period, above all Germany.

Note that catching-up vis-à-vis the United States did not yet mean reaching the United States in terms of economic performance. FIGURE 99 summarizes an index of labour productivity in manufacturing for a sample of countries at different historical points of time, covering 1870, 1950, 1973 and 1989, with the UK productivity level set at 100. The table shows that, by 1973, the big productivity gap between the United States and Europe had been narrowed down, but not quite disappeared. Even by 1989, a narrowed gap survived. Sure enough, any attempt at comparing productivity levels in manufacturing over such very long time spans must be interpreted with great caution. However, the sheer magnitude of the long surviving transatlantic gap does suggest that the process of convergence was far from completed by 1973.

How, after a devastating war, was Europe put back on a track of growth and convergence? How did the resurgence and re-integration of Europe work? Right from the start, all European countries faced formidable obstacles to re-entering world markets. The most important one was clearly a lack of convertible currency, what at that time was called the “Dollar-shortage”. In the late 1940s and early 1950s, immediate post-war reforms had brought some recovery of production and basically the return to pre-war production levels in most European countries (recall FIGURE 94). In a true sense, it was the time of reconstruction and repairs. However, to go beyond that, a genuine re-establishment of a broader trade network across Europe was necessary. For this purpose, however, an accepted money for transaction purposes was needed. In the prevailing circumstances, this money could only be the US-Dollar. With currency convertibility not yet established, some way had to be found to re-build a workable intra-European system of payments.

FIGURE 100 shows graphically where the major problem lay. It indicates the degree of over-valuation of major European currencies as measured by the deviation of the official Dollar parity from the exchange rate as traded in the free Zurich market. At no time until 1953 did the Pound Sterling, the French Franc and the (still young) German Mark reach a market value that was equal or higher than the official parity, and at some times the remaining gaps were substantial. Given this starting-point, a hefty devaluation vis-à-vis the Dollar might have been considered. However, in the actual politics of the time, hardly anybody regarded this as a real option. After all, there was a widespread feeling that a devaluation-induced sharp rise of imported food prices might ignite social unrest in view of post-war scarcities and widespread misery. In addition, notably in Germany, the reputation of the new currency was at stake, and the risks of a wage-price-spiral were regarded as prohibitively high in case of devaluation.

In view of this European dilemma, the Americans stepped in with a major support programme, which aimed at a fast and smooth integration of the European economies in world markets. To be sure, some humanitarian aid had been provided by the United States right after the war. However, from 1947 onwards, the aid was converted into the fully-fledged "European Recovery Program" (ERP), which became better known as Marshall-Plan.

Institutionally, the ERP consisted of the foundation of a European Cooperation Administration (ECA) in Paris, which was to steer and supervise the programme from the American side. Parallel to this, as the European counterpart of the programme, the Organisation for European Economic Cooperation (OEEC) was created, also with seat in Paris. Note that, once its post-war task had been successfully completed by the late 1950s, the OEEC was transformed into the broader Organisation for Economic Cooperation and Development (OECD), which has survived and flourished up to the present as a major political, economic and research arm of the industrialized part of the world.

Economically, the Marshall-Plan granted help for a variety of purposes, mostly in the form of grants (88 %) and to a much smaller extent of loans (12 %). As to the purposes of aid, almost one half focused on scarce raw materials, roughly one third on food and the rest on machines, vehicles and miscellaneous investment goods that were urgently needed. Generally, the aid was tightly supervised by the ECA so that the likelihood of a misdirecting of funds could be minimised.

FIGURE 101 gives a statistical overview of the volume of the funds and the regional allocation over European countries in the relevant period 1948-53. The table shows that, in eco-

nomic terms, the volume of the aid was not so big as one might expect in view of the Marshall fund's honorary place in post-war history. After all, it amounted to no more than an average of 2.1 % of the Gross National Product of all recipient countries. In Germany, the respective share was even below average (1.4 %). Hence there is hardly a statistical basis for attributing a fundamental economic importance to the Marshall plan for Europe's recovery (and in particular for the recovery of Germany). Be that as it may, much qualitative and anecdotic evidence indicates that Marshall Plan aid did help to remove important bottlenecks in industrial production lines that were to gain economic importance in the years to come.

The major significance of Marshall Plan aid was anyway a political one, with important economic consequences in the long run. With the Marshall Plan, the United States combined a friendly pressure towards opening up European markets and pushing even reluctant countries such as, e. g., France into a more positive attitude towards liberalisation of trade and payments. In a way, "big brother" America signalled to Europe and the world that it stood ready as a friend and supporter of a liberal trade order within Europe and beyond, thus shouldering the responsibility of a benevolent hegemon who had a major interest in a constructive development. In this sense, the big lesson of the interwar period had been learnt, and the Marshall plan became the visible symbol of it.

In fact, it was this investment in liberalisation that quite obviously delivered high political and economic returns. To further the intra-European financing of trade, the European Payments Union (EPU) was founded early on – as a sort of clearing house that provided quasi-automatic balance of payments loans between countries with surpluses and deficits. In turn, trade liberalisation proceeded fast: by the second half of the 1950s, intra-OEEC trade was largely freed from all quantitative import restrictions that had still existed on a broad scale of products in the early post-war period.

To be sure, all this was part of a larger move towards free trade in the framework of the General Agreement on Tariffs and Trade (GATT), which had been concluded in 1947 by 23 countries. It was to become the first major multilateral trade agreement of the world that turned out very successful in providing a framework for successive rounds of tariff reductions and for establishing principles of non-discrimination like the most favoured nations clause (MFN), which had existed in bilateral agreements of the 19th century, but not as a generally accepted element of "good conduct". For the first time in history, the goal of free trade became a major aim of a multilateral agreement, and ever more countries joined that agreement over the coming decades. Eventually, in 1994, the GATT was replaced by the World Trade Organisation (WTO), which took over even wider responsibilities than its contractual predecessor in

the fields of trade dispute settlement, service trade and the protection of intellectual property rights. Nowadays, with more than 150 member countries, the WTO has become the uncontested guardian of and forum for world-wide trade issues, though also a major target of Globalization critics.

In the 1950s and 1960s, the GATT was eminently successful in reducing the world-wide tariff rates in successive so-called “trade rounds”. And it served as a framework in which countries with a pronounced protectionist tradition could bind themselves into agreements that proved very advantageous for their own export industries in the long run. In this respect, the paradigmatic case is Germany. A member of the GATT since 1950, it started into the trade rounds with its tariff structure still going back to the early 20th century (the so-called Bülow tariffs), but soon dismantling them in the subsequent multilateral negotiations. In the end, Germany became a remarkably trade-oriented country, with a national volume of exports that has been the largest in the world in recent decades.

3.3.1.3 The Surge of Intra-Industry Trade

The resurgence of Europe in trade was strongly linked to the foundation of the European Economic Community (EEC) in 1957. Once the first major steps towards re-integration in world markets had been achieved, the six founding countries of the EEC – France, Italy and Germany as well as the Netherlands, Belgium and Luxemburg – formed a customs union with the long-term aim of a common market with free movement of goods, services, labour and capital. As a reaction to the birth of the EEC, the European Free Trade Association (EFTA) was founded in 1960 – as a looser community of non-EEC-countries that established a free-trade area (though not a customs union with a common trade policy).

Thus, for a while, western and central Europe was formally split into two trade zones. And it is remarkable that the sustained expansion of intra-European trade between 1960 and 1973 did strongly reflect this split. This can nicely be seen from the volume and structure of the fast growing West German foreign trade in the 1960s and early 1970s (though other countries’ trade could be taken to show basically the same). FIGURE 102 provides numbers on West German trade values and volumes from 1960 to 1973. Clearly, the growth was remarkably fast and sustained, with growth rates hovering around 10 % p. a. all over the period. A glance over West Germany’s regional pattern of trade, however, does reveal that trade growth was disproportionately fast within the EC (the European Community as it was later called). FIGURE 103 presents numbers: the share of West German EC-trade grew from below 30 % in 1960 to 40-47 % by the early 1970s, whereas the share of Trade with EFTA-

countries shrank markedly in the same period (on the import side: from about 20 to 12 %, and on the export side: from about 28 to 23 %).

To be sure, this divisive trend in Europe did not last. With the EC-enlargement in 1973, which turned the United Kingdom, Ireland and Denmark from major EFTA- into EC-members, much of the trade division of Europe was removed and gradually replaced by a more rational trade structure. With further waves of EC-enlargement in 1981, 1986 and 1995 and free trade agreements between the rump-EFTA and the European Union (EU) as it is called since 1992, the remaining obstacles to a liberalized trade structure within western and central Europe were removed.

By and large, the European trade integration that was begun by the EEC and then continued by the EC and the EU was a remarkable success. However, the success was strongly concentrated on the fast growing segment of intra-industry trade in manufacturing. FIGURE 104 delivers some evidence on this point for the case of Germany: between 1960 and 1980, the share of intra-industry in total trade rose in manufacturing from 62 to 81 % for intra-EEC-trade and from almost 60 to 72 % for trade with EFTA. Even for trade with the rest of the world, the trend towards intra-industry trade is unmistakable, with trade shares growing from around 50 % in 1960 to around 70 % in 1980. Apparently, an increasingly advanced industrialized country like West Germany became an ever more integrated part of the world-wide trade in manufacturing goods at all stages of production, from intermediate to final goods.

Sure enough, this was a general trend in Europe at that time, which is an integral part of the catching-up process. Much more than in the 19th century, the trade growth in the second half of the 20th century came down to an ever more intensifying division of labour within narrow branches of manufacturing. It is a trade that can best be explained in terms of models that combine the working of economies of scale on the supply side with a differentiated structure of tastes and technical needs on the demand side. It is not anymore the kind of inter-industry trade that can be properly described by classical or neoclassical trade models of the Ricardo- or Heckscher-Ohlin type, which focus on international differences of factor productivities and factor endowments.

Given the great success of European integration in the process of catching-up to the United States in the 1950s up to the 1970s, the focus on intra-industry trade and the resulting specialisation patterns can only be viewed as a most viable strategy of growth and development. However, it does have a dark side, and that is agricultural protectionism. From the start, the European Community set up a so-called “common agricultural policy” (CAP), which involved

a large-scale subsidisation of dairy and grain farming. In view of low world market prices for agricultural products, the CAP could only be upheld with massive trade distortions and behind high walls of trade barriers against agricultural imports, to the detriment notably of developing countries. Hence the bias towards intra-industry trade growth was to some extent the consequence of deliberately keeping out those products that constitute the classical type of trade between agrarian and industrialized countries. In this sense, the European Community followed the tradition that had been firmly established in continental in the last decades before World War I, as we have seen in section 3.1.

3.3.1.4 The Bretton Woods System

As early as July 1944, delegations from 44 countries gathered in Bretton Woods, New Hampshire. They laid the contractual ground for an international agreement that, after the war, should become the basic monetary framework for cross-border exchanges of goods and services. The agreement was later called the Bretton Woods System, and until its breakdown in 1973, it did in fact serve exactly in the role that its intellectual fathers had in mind. One of them was John Maynard Keynes, who is generally considered as the main designer of Bretton Woods. In fact, the architecture of the system carries many features of a Keynesian outlook that accepts external stability as a useful and important goal, but allows some pragmatic deviation from it whenever internal macroeconomic disequilibria become unbearable as had been the case in the Great Depression.

In essence, the Bretton Woods System was constructed as a so-called Gold Exchange Standard, with gold and the US-Dollar serving as reserves. The Dollar-price of gold was fixed at 35 \$ an ounce, and all currencies in the system had a fixed Dollar price, thus in effect leading to an overall system of fixed exchange rate. All central banks (except the American one, the Fed) were obliged to intervene with purchases or sales of Dollars and/or gold, once the market exchange rate of their currencies threatened to deviate from the fixed parity. Hence, in principle, no independent monetary policy was possible.

In this sense, external balance had priority over internal considerations. But the Bretton Woods system did allow some flexibility:

- A pool of reserves was created – and commonly financed with gold and US-Dollar reserves – that stood ready to grant stabilization loans in case of temporary balance of payments crises of countries. This pool was to be the “International Monetary Fund” (IMF), and the granted loans were called the IMF-lending facilities.

- In case of a so-called fundamental disequilibrium, parities of individual currencies in terms of US-Dollars (but not the Dollar gold price itself) could be adjusted, thus allowing a country to use a kind of emergency exit to avoid extremely costly downward pressures on internal absorption.

Hence, clearly, the Bretton Woods system was more flexible than the traditional Gold Standard. More than that: as its aim was merely convertibility on current account transactions and thus explicitly allowed controls on short-term capital flows, it could be expected to work under less external pressures of speculative financial movements than was the case with the Gold Standard in the late 1920s (recall in this respect the dismal German experience!). In fact, by the late 1950s, most countries had established convertibility on current account transactions, and Japan followed suit in 1964. However, the removal of capital controls remained mostly confined to countries with a strong liberal tradition in financial markets (like Britain) or very comfortable external surpluses over longer periods of time (like Germany).

So much for the basic principles of the system. How did it work? The answer is: it worked very well for quite a long (and important) time, but then it broke down much faster than anyone of its advocates might have believed. The years of success were those when private short-term capital flows were on a relatively small scale and American fiscal and monetary policy strictly stability-oriented. In a sense, these were the years of “good weather”: continental Europe being on its path towards liberalisation and fast growth, but its capital markets still in their infancy; and the United States really adopting the role of the United Kingdom in the pre-World War I-Gold Standard as a haven of stability and of current account surplus generation. Remarkably, some economists like Robert Triffin warned that, in the longer run, the system might be undermined in its credibility due to the much faster economic growth in Europe, which also meant that European central banks were to accumulate ever more reserves in gold and Dollars. Once the sum of these reserves would dwarf the American ones, the willingness of Europeans to hold Dollars might dwindle, and so the confidence in the system. But these were, for the time being, academic arguments that did not by themselves make the weather worse.

By the mid 1960s, however, the weather turned worse by itself. The first major crack in the system became visible when Britain faced record trade deficits so that a devaluation of the British Pound turned out to be inevitable by November 1967, after massive speculative attacks. Another crack emerged when, in 1969, a combined devaluation of the French Franc and revaluation of the German Mark took place – and thus made clear to the markets that

notably Germany's rising current account surpluses could not be reduced without a change of parities.

More than anything else, however, it was a fundamental shift of macroeconomic policies in the United States that destroyed the basis of the system. In 1965-8, President Johnson stepped up military expenditures for the Vietnam war and at the same time put on an ambitious social programme called "Great Society". All this happened without a tax increase so that internal absorption rose rapidly. The consequences can be recognized from FIGURE 105: government purchases soared with high growth rates, the inflation rate increased substantially, the traditional current account surplus diminished drastically and, maybe worst of all, the money supply expanded rapidly, thus indicating a partial monetisation of public indebtedness. As a consequence, massive speculation against the Dollar in the London gold market in 1967/8 led the United States to adopt a so-called two-tier system, which allowed for a flexible gold price in US-Dollars in the free market, but not for official transactions. With the benefit of hindsight, this step might be considered as the beginning of the end of Bretton Woods because the link between free markets and official parities was severed for the first time. The system thus lost a major market anchor and thus quite a bit of its credibility.

By 1970, a beginning recession in the United States made it clear that, economically, a Dollar devaluation was inevitable. Technically, this could only be arranged within the Bretton Woods rules by a revaluation of all other currencies, which was not in the interest of most countries involved and therefore required difficult multilateral negotiations. As these were not successful, the American president Richard Nixon announced the end of US-gold sales and a 10%-tax on all imports, apparently as a measure to push the other countries back to the negotiation table. In that, he succeeded, and in the Smithsonian Agreement of December 1971, an 8 % devaluation of the US-Dollar was granted in return for the removal of the import tax. However, the widely hailed terms of the Smithsonian Agreement saved the Bretton Woods System for just another 14 months. As the US-current account turned into deficit and other macroeconomic statistics indicated ever less stability of the American economy, massive speculative attacks on the Dollar led to the closing down of foreign exchange markets at the beginning of March 1973 and the transition to free floating on march 19, 1973.

While the story of the Bretton Woods breakdown has quite a few dramatic features, the deep causes can be easily read off the standard statistics of all countries involved. FIGURE 106 shows the inflation rates of Britain, France, Germany and Italy in the critical period from 1966 to 1972. It is quite obvious that the sharp acceleration of inflation from 1969 to 1972 was a classical incident of "imported inflation": as American fiscal and monetary expansion turned

once again inflationary, European central banks were forced to expand their money supplies by purchasing Dollars, and thus fuelling domestic inflation. Notably in a country like Germany, whose population had lived through one hyperinflation and two currency reforms within two generations, that kind of policy was unpopular and unsustainable in the long run. FIGURE 107 shows how strong the pressure on the Bundesbank, the German central bank, turned out to be at the peak of the crisis from 1970 to 1972: money supply growth accelerated to double-digit rates, and the Dollar reserves swelled to record levels.

The political lesson from all this is clear: once the United States were not ready anymore to accept domestic sacrifices for external stability, the Bretton Woods System was doomed to fail. It is remarkable that, since 1973, there has been no further attempt to re-establish a worldwide system of fixed exchange rates. And it is this very fact that indicates that the time for such a system was over precisely because the most important nation, the United States, was not interested anymore to shoulder any big responsibility that is invariably linked in such a system to the role of the biggest country and economy.

To be sure, the end of a worldwide system of fixed exchange rates did not mean the disappearance of any such system. In fact, the monetary history of Europe after Bretton Woods is a remarkable example to the contrary. A few years after the breakdown of Bretton Woods, the European Monetary System (EMS) was established – on the initiative of the French President Valéry Giscard d'Estaing and the German Chancellor Helmut Schmidt. The EMS was a fixed exchange rate system that allowed a certain fluctuation of currencies in a narrow band of 2.25 % of an assigned par value, with some economically weaker countries being allowed a larger band of 6 %.

The historical record of the EMS is a mixed one. On the negative side, there were quite a few currency realignments and also major crises, the last one in 1993: German unification had led to a combination of a massively expansionary fiscal policy at a tight monetary stance of the Bundesbank, with the consequence of high interest rates in Germany, which led to a revaluation pressure on the Mark (and correspondingly devaluation pressures on other currencies). That incident even led to a widening of the bandwidth to 15 %, which was not at all in the convergence spirit of the EMS. On the positive side, however, capital controls were removed by 1987 in all countries adhering to the EMS. And, in the Maastricht Treaty of 1991, the path towards a common European currency was laid, with the European Central Bank (ECB) as an independent central bank providing one currency for “Euroland” by the early 21st century, which then in fact happened.

Was this a success? Politically, it certainly was as it showed a remarkable determination on the side of the politicians in charge to carry out a project that was not particularly popular to begin with and received quite sceptically by many economists and other experts. Economically, it remains an open question whether “Euroland” (see the dark areas in FIGURE 108) qualifies as some sort of optimum currency area, with a high degree of factor mobility to compensate for the lack of adjustable parities, and with a high degree of trade integration. It remains to be seen in the years to come whether, in this sense, Europe can ever more resemble the United States, which cover a very large territory, but have been a successful currency area for a very long time.

In one respect, however, the EMS and its eventual transformation into a common currency area has been a remarkable success: in terms of macroeconomic convergence. True enough, the standard goals of convergence in the Maastricht treaty – a maximum of 3 % public deficit as a share of GDP and a maximum of 60 % ratio of debt to GDP – have been frequently violated, though a trend in the right direction is visible. Be that as it may, it is the convergence in inflation towards a low common rate that has been spectacular. FIGURE 109 shows annual inflation rates relative to the German rate, from 1978 to 2002. The picture is striking: From a vast array of different rates in the late 1970s, all countries came down to the (very low) German rate. A similar picture could be shown for interest rates, which indicates that – even without deficit and debt criteria properly met – financial markets appear to honour the path towards a strictly stability oriented policy of an independent central bank. It is likely that this leads to relatively low capital costs in the longer run, in a “Euroland” that explicitly renounces on capital controls and price inflation. Only a resurgence of inflationary pressures as they are in fact visible at present might endanger this positive outlook.

3.3.2 Oil Crises and Their Aftermath

With the benefit of hindsight, the year 1973 marks a quite dramatic change in world economic history. One major event was the breakdown of the Bretton Woods system in March. The second one was the first oil price hike in October. The story is simple enough: in that month, a war broke out between Israel and the Arab states, and the Organisation of Petroleum Exporting Countries (OPEC) staged an embargo on oil shipments to the most Israel-friendly countries, the United States and the Netherlands. As the OPEC was widely considered as an effective cartel with a strong influence on world prices of oil, speculative waves further pushed up the price of oil, which reached about 12 US-Dollars (up from 3 US-Dollars at the beginning of the price hike). As usual when oil becomes more expensive, other raw material prices followed suit.

The new constellation hit virtually all industrialized countries hard. It worked like a ubiquitous price shock and cost push that came unexpectedly, although the Club of Rome, an independent think tank, had described a frightening scenario of oil reserve depletion in a report published in 1972. That report was widely read, but probably more as a piece of apocalyptic fiction than a harbinger of a sharp oil price rise. With the benefit of hindsight, it is clear that a long period of high energy prices was ushered in, with another oil price shock coming after the Iranian crisis in 1979/80 and raw material prices staying high until about 1986.

In macroeconomic terms, a new age began (see FIGURE 110). Economic growth slacked everywhere: compared to the 1950s and 1960s, the per-capita trend growth rates looked altogether disappointing, though they were in fact just a return to normality if one takes a proper long-run view (see, e. g. FIGURE 93). Unemployment rates jumped up in industrialized countries, first to levels of about 5 % and then, at the beginning of the 1980s, to levels of 7-10% (see again FIGURE 110). And inflation soared up to average annual rates of 8-11 per cent in the period 1973-82.

How was all this possible? To explain it, one has to focus on two major effects that hit the western industrialized economies in an almost parallel fashion. They are most clearly visible for the case of Germany, the country that had enjoyed a spectacular economic resurgence in the two decades before 1973:

- The oil price shocks led to a drastic worsening of Germany's terms of trade (see FIGURE 111). As imported raw material prices rose sharply in two big hikes (1973-75 and 1979-81), the terms of trade worsened accordingly, and this worked like a big cost push for German manufacturing industry, which had been so successful in expanding in world markets in the decades before 1973.
- In a state of extremely tight labour markets with unemployment rates below 1 % in some boom times (see FIGURE 112), the early 1970s became a period of exceptionally fast nominal wage growth that by far surpassed the rate of inflation and productivity advance (see FIGURE 113). Hence, on top of oil prices, there was another cost push to be digested by manufacturing industry, a situation that had practically never existed before, even not in the troubled 1920s.

Given this exceptional combination of circumstances, it is not quite so surprising that, apart from some critical observers, most of the economic and political establishment did not really understand what was going on. Once the cost push led into a major recession in 1974/5, the

first one since the Great Depression in 1930/2, they interpreted the situation just as another deep recession of the style they knew from the early 1930s.

FIGURE 114 tries to visualize how politicians and central bankers committed a major “identification error” in 1973/75. In terms of a standard short-run scheme of aggregate demand (AD) and aggregate supply (AS) in output/price-level-space, they thought to see an aggregate demand shock, i. e. a standard leftward shift of the AD-curve, just as had happened in the Great Depression (FIGURE 114, left-hand side). They did not recognize that what had happened was in fact an aggregate supply shock, i. e. a shift of the AS-curve to the left due to sharply rising prices of raw materials and labour (FIGURE 114, right-hand side). Clearly, what could be observed was not “deflation”, with output and prices falling, but “stagflation”, with output falling and prices rising.

This misinterpretation of the facts had serious political consequences. In the recession 1974/5, fiscal and monetary policies turned expansionary in virtually all industrialized countries so as to counter the output contraction. A leftward shift of the AS-curve was thus answered by a rightward shift of the AD-curve, leading to a further fuelling of inflation (see FIGURE 115, left-hand side). Not before the second stagflation shock in the early 1980s did policy react differently. Then came a tough anti-inflationary stance, i. e. the leftward shift of the AS-curve was answered by a leftward-shift of the AD-curve due to a restrictive monetary policy (see FIGURE 115, right-hand side). In the end, this policy was successful in fighting inflation (see FIGURE 116 for the case of Germany), but in doing so, it exacerbated unemployment.

So much, in a very stylized form, about the basic macroeconomics of the period under consideration. To be sure, the shift of policies between the two oil price crises and recessions came down to a kind of intellectual and political counterrevolution against the Keynesian consensus that had dominated academics and policy-making from the mid 1930s to the mid 1970s. In a sense, the lesson that had been learned after the Great Depression had to be completely re-written in the light of the stagflation experience. Among academic economists, this led to the rise of monetarism and the so-called rational expectations school, which since then is a major ingredient of standard macroeconomic knowledge and teaching.

Beside these new insights about the nature of the business cycle, a new empirical phenomenon emerged that had not been observable since the interwar period: persistent mass unemployment. After the recession 1974/5 was over and capacity utilization in industry approached normal levels again, it became soon clear that unemployment did not return to the

low pre-recession level in virtually all industrialized countries. That led to political calls for further demand expansion notably in Germany and Japan, who were running large current account surpluses and appeared to have the macroeconomic leeway for an expansionary monetary and fiscal policy, which in fact came about in 1978/9. Even then, however, unemployment remained stubbornly high by previous standards, and on top of all came the second oil price hike in 1979/81, which fuelled inflation once again. This ushered in a major policy shift towards monetary restriction. In October 1979, the newly elected chairman of the Federal Reserve System Paul Volcker credibly announced a tough anti-inflationary stance. Quite soon, other central banks followed suit, not least because price inflation was reaching peak rates again everywhere. The recession that followed was long and sustained, and it led to a further rise of unemployment.

Up to this point, the two oil crises and recessions had raised unemployment on both sides of the Atlantic to new peaks that had not been seen since the interwar period. From the early 1980s on, however, developments in the United States and Continental Europe began to diverge, with Britain being a kind of case in between: While the world-wide cyclical recovery from 1983 onwards gradually pulled the American unemployment rate back down to the level of about 5 % that had been standard in the 1950s and 1960s, it did not have the same effect in Europe. In the vast majority of countries on the old continent, unemployment remained very high all over the 1980s and the 1990s. In fact, up to the very present, most countries in Europe have not returned to the degree of full employment that they had realized in the 1960s and early 1970s. Apparently, the aggregate supply shocks that had hit product and labour markets led to a permanent rise of the non-cyclical unemployment rate, which is determined by long-term structural forces. In econometric terms, what was observed in European labour markets could be characterized as “path-dependent”: the crises passed, but they left long-term traces that prevented a return to the shining labour market performance of earlier times.

What were these traces? An enormous amount of research has been done on the issue of why the “natural rate of unemployment” or how ever one may call it did change in continental Europe, but apparently not in the United States (and to a much lesser degree in Britain). Without going into any details of this highly inspiring debate among macro- and labour economists, we may clarify one central point of an explanation by juxtaposing the experience of two big countries: Germany and the United States. As is well known, Germany has a relatively generous system of unemployment insurance, which in the relevant period provided any laid-off worker with a claim to roughly 60 % of his terminal wage as unemployment benefits that were granted indefinitely. To be sure, many other European countries had similar sys-

tems, though in recent years some of these (and also the German one) have been made much less generous (in Germany through the so-called Hartz-IV reforms). In the United States, no such system ever existed so that, after an unemployment spell of six months, virtually all unemployed persons fall back on (modest) welfare support.

Given this institutional difference between the United States and Germany, the labour markets of the two countries were bound to react very differently to the drastic fall of manufacturing jobs that happened during the crises of 1973/5 and 1981/3. Typically, the crises led everywhere to a big net loss of well-paid blue-collar jobs. In the United States, the laid-off workers had to search for new employment fast and intensively as they could not rely on any indefinite payment of an unemployment benefit that was a relatively high share of their previous wage. They had to accept even low-pay service jobs, and they did. As a consequence, the long-term effect of the crisis was a rise in income inequality, but not in unemployment. Once the crisis was over, all had jobs again, but many at lower wages than before. In Germany laid-off industrial workers remained jobless as the unemployment support system allowed them to restrict their job search to relatively well-paid segments. In fact, they became the core of the many long-term unemployed persons, who had enormous difficulties to be reintegrated into the labour market due to a lack of skills that might qualify them for better-paid service jobs.

With some courageous simplification, the structural change away from industry and its labour market effects can be demonstrated by juxtaposing the unemployment rate (u) and the share of non-industrial employment (NIS) in West Germany and the United States for the period 1950-90 (see FIGURE 117). The message is straightforward: for West Germany, the sharp rise of the share of NIS between 1973 and 1990, which involved a net loss of 1.4 million industrial jobs, went along with a sharp rise of the unemployment rate u . True enough, there were still some cyclical fluctuations of u along the upward trend path, but they were minor compared to the power of the upward trend. For the United States, the picture looks totally different: between 1973 and 1990, NIS goes up in trend, but the unemployment rate u follows a pronounced cyclical path with no long-term trend to be recognized. By the late 1980s, labour markets in the United States were roughly back at where they had been in the early 1960s, with the turbulent period in between being not more than a temporary aberration.

Note that the comparative experience of Europe and the United States contains a general lesson of how the construction principles of an unemployment support system and a welfare state work under different external conditions. Recall that, in section 3.1.1, we interpreted “the birth of unemployment” and “the rise of the welfare” as a consequence of industrializa-

tion. Remarkably enough, the de-industrialisation that was the consequence of the supply side shocks in the 1970s and 1980s clarifies that it can matter very much for the labour market how the welfare state is constructed. This is illustrated in FIGURE 118 where two types of structural change are graphically described for a world with two sectors of which sector 1 – call it industry – pays high wages and sector 2 – call it services (or agriculture) – pays low wages at any level of employment. Note that the graphs in FIGURE 118 are drawn in a way that the total labour force L of the economy can move between sectors 1 and 2, but that there remains a differential of wages (w) between the sectors in the relevant range. Note also that this is a realistic assumption, at least for unskilled labour whose marginal productivity tends to be much higher in industry as they mainly rely on blue collar physical work.

Let us now assume that, due to external market forces – say, a fast integration of industry into world markets as in the 1950s and 1960s –, labour demand increases in sector 1 and decreases in sector 2 at given wages w^{*1} and w^{*2} . Then workers will be laid off in sector 2 (at wage w^{*2}), but will voluntarily move from sector 2 to sector 1 because they are better paid there (at wage w^{*1}). Note that this structural change (“industrialization”) runs smoothly even with a generous system of unemployment support that leads to sectoral reservation wages (denoted by superscripts “r”) that are not much below the actual wage paid in the respective sector. After all, the industrial wage is much higher than the service sector wage, and hence also much higher than the service sector reservation wage. Hence employment (N) increases in sector 1 and decreases sector 2, with no increase in unemployment.

The situation is totally different when structural change works the other way round (as happened in the 1970s and 1980s): from sector 1 (“industry”) to sector 2 (“services”). This is illustrated in the second part of FIGURE 118 with a leftward shift of the two labour demand curves. Due to high unemployment benefits, laid-off industrial workers have a reservation wage that is higher than the wage that is actually paid in the service sector. Hence they will continue to search for industrial jobs and become long-term unemployed, as a consequence of “de-industrialization”. The shrinkage of industry has now a ratchet effect on unemployment. Note that this would not happen if the reservation wage were uniformly low for both sectors due to, say, a welfare system (as the American one) that does not differentiate in the long run according to the employment history of the individual unemployed person.

So much for the long-run consequences of the two supply side crises for European in contrast to American labour markets. In general macroeconomic terms, the 1980s were a time of very different policy mixes in the United States and Europe. While monetary policy was stability oriented on both sides of the Atlantic, fiscal policy was for a long time much more ex-

pansionary in the United States than in Europe. After the presidential election in 1980, American military expenditures were vastly increased and taxes cut so that a large “twin deficit” emerged in the public budget and the current account. This new policy mix was later called “Reaganomics”, after the newly elected President who was responsible for the fiscal part of it. As the high absorption of the American economy was accompanied by a restrictive monetary policy stance, American interest rates soared, and so did the external value of the US-Dollar, which rose by roughly 50 % in real terms between 1981 and 1985 (see FIGURE 119).

Not before the middle of the decade did this macroeconomic configuration begin to change. It did so for a mixture of political and economic reasons. Politically, domestic protectionist pressures became virulent in the United States as American manufacturing massively suffered from the high valuation of the US-Dollar. In turn, governments and central banks in Europe began to regard a slight switch towards expansion as much less risky as in earlier years. After all, by 1985/6, the victory over price inflation became apparent, oil prices began to drop and the European economies were still in a stage of relatively sluggish growth. All this came down to a transatlantic consensus on signalling to financial markets that governments and central banks stood ready to commit themselves to preventing a further rise of the US-Dollar. This was ostentatiously done in a so-called G5-meeting in the New York Plaza Hotel, and the markets apparently took it as a credible message. In the post-Plaza-period, the Dollar devalued rapidly and stayed low until the mid 1990s (see FIGURE 119).

Note that the Plaza agreement, as it was called, was the first major move away from free floating since the breakdown of the Bretton Woods System. Whether it worked by itself or whether the macroeconomic constellation was gradually changing anyway can remain open at this point. Sure enough, it fell into a period of a tightening of the European Monetary System, which promised a relatively easy convergence in due course. However, this honeymoon did not last: in 1990 a recession hit the United States and German unification led to a massive expansionary fiscal policy in Germany, which had some parallels to the American “Reaganomics” a decade earlier. As at that time, the policy mix of fiscal expansion and monetary stability led to a sharp revaluation of the currency, which ushered in the last big crisis in the EMS and a further decline of the Dollar. Again, domestic considerations had prevailed, and it is fair to say that any worldwide system of fixed exchange rates, had it still existed, would have had a very bad time under the resulting pressures.