1
Center and Periphery

About a year ago I more or less suddenly realized that I have spent my whole professional life as an international economist thinking and writing about economic geography, without being aware of it.

By "economic geography" I mean "the location of production in space"; that is, that branch of economics that worries about where things happen in relation to one another. It's not worth trying to define my subject more exactly than that—you'll see better what I mean once I start describing models. Most of regional economics, and some but not all of urban economics, is about economic geography in the sense I have in mind.

If you had never looked at the theory of international trade, you might have supposed that international economics would also be largely treated as a special case of economic geography, one in which borders and the .actions of sovereign governments play a special role in shaping the location of production. What I will argue in these lectures is that that is how international economics ought to be done, at least part of the time. But

it is almost never the way that it is done at present. Instead, the analysis of international trade makes virtually no use of insights from economic geography or location theory. We nor-really model countries as dimensionless points within which factors of production can be instantly and costlessly moved from one activity to another, and even trade among countries is usually given a sort of spaceless representation in which transport costs are zero for all goods that can be traded.

There is nothing wrong with simplifying assumptions—on the contrary, it is only through strategic simplification that we can hope to make any sense of the buzzing complexity of the real world. The particular simplifying assumptions of conventional trade theory have led to an impressive and very useful intellectual construct. For some purposes it does no harm to ignore the fact that countries are not points and that some pairs of countries are much closer than others—that California is farther from New York than any place in the European Community is from any place else, or that London and Paris are much closer to each other than are New York and Chicago, or for that matter that Canada is essentially closer to the United States than it is to itself.

Yet the tendency of international economists to turn a blind eye to the fact that countries both occupy and exist in space—a tendency so deeply entrenched that we rarely even realize we are doing it—has, I would submit, had some serious costs. These lie not so much in lack of realism—all economic analysis is more or less unrealistic—as in the exclusion of important

issues and, above all, of important sources of evidence. As I hope I will be able to show, one of the best ways to understand how the international economy works is to start by looking at what happens *inside* nations. If we want to understand differences in national growth rates, a good place to start is by examining differences in regional growth; if we want to understand international specialization, a good place to start is with local specialization. The data will be better and pose fewer problems of compatibility, and the underlying economic forces will be less distorted by government policies.

The decision by international economists to ignore the fact that they are doing geography wouldn't matter so much if someone else were busy exploiting the facts and insights that can come from looking at localization and trade within countries. Unfortunately, nobody is. That is, of course, an unfair statement. There are excellent economic geographers out there, as well as urban and regional economists who worry about geographical issues. For reasons that I will discuss in a moment, however, these people are almost uniformly peripheral to the economics profession. International economics is a flagship field: no serious economics department Can get by without at least one international trade expert and without offering international economics as a field for its graduate students. By contrast, regional and even urban economics are given far less priority. And economic geographers proper are. almost never found in economics departments, or even talking to economists; at best they are in urban studies departments, more usually in geography departments. They may do excel

lent work, but it does not inform or influence the economics profession.

There are good reasons why this has happened and equally good reasons why it should change. Before I begin to present my own ideas, I want to talk briefly about why international economists don't acknowledge that they are doing geography—and why they should.

Geography: Why Not and Why

The neglect of spatial issues in economics arises for the most part from one simple problem: how to think about market structure. Essentially, to say anything useful or interesting about the location of economic activity in space, it is necessary to get away from the constant-returns, perfect-competition approach that still dominates most economic analysis. As long as economists lacked the analytical tools to think rigorously about increasing returns and imperfect competition, the study of economic geography was condemned to lie outside the mainstream of the profession. Indeed, as standards of rigor in economics have risen over time, the study of location has been pushed further and further into the intellectual periphery.[1]

[1.] A major exception is urban economics, where there is a strong modeling tradition that informs a large body of empirical work Henderson (1974,1988), in particular, has developed a very persuasive framework for analyzing the evolution of an urban system and has provided extensive empirical evidence in support. l think it is fair to say, however, that international economists have largely ignored or been unaware of this body of work

Not all students of economic geography have understood this. Much of the literature on industrial location, in particular, has ignored the issue of market structure and instead been obsessed with geometry—with the shape of market areas on an idealized landscape, or with the optimal siting of facilities given markets and resources—while paying little or no attention to the problem of modeling markets. This is, to my mind, doing things in the wrong order, worrying about the details of a secondary problem before making progress on the main issue.

Step back and ask, what is the most striking feature of the geography of economic activity? The short answer is surely *concentration. Think* of the United States: most of the population of a huge, fertile country lives along parts of two coasts and the Great Lakes; within these belts, population is further concentrated in a relative handful of densely populated urban areas. As I will document in the next lecture, these urban areas in turn are highly specialized, so that production in many industries is remarkably concentrated in space.

This geographic concentration of production is dear evidence of the pervasive influence of some kind of increasing returns. And there is the problem. Increasing returns are simply harder to model than constant or diminishing returns. If the increasing returns are purely external to firms, we can still use the tools of competitive analysis; but external economies turn out to be both analytically awkward and empirically elusive. If the

increasing returns are internal to firms, we are faced with the necessity of modeling imperfect competition.

Economics tends, understandably, to follow the line of least mathematical resistance. We like to explain the world in terms of forces that we know how to model, not in terms of those we don't. In international economics, what this meant from Ricardo until the 1980s was an almost exclusive emphasis on comparative advantage, rather than increasing returns, as an explanation for trade.[2] The point was that comparative advantage could be modeled using models that assumed constant returns and competition, which were the tools at hand. The profession simply put those aspects of international trade that could not be modeled that way on one side.

Unfortunately, the evident importance of increasing returns in economic geography is so great that this understandable impulse to focus on what we know how to deal with has led to an avoidance of the subject as a whole. After 1940, in particular, as the expected level of rigor in economic discussion steadily rose, economic geography was simply submerged.

But times have changed. During the 1970s there was a new wave of theory in industrial organization, which provided the economics profession with a menu of models of imperfect

2. For those who worry about definitions, by comparative advantage I mean the general idea that countries trade in order to take advantage of their differences. The increasing returns approach asserts instead that countries trade because there are inherent advantages to specialization, even for initially similar countries.

competition. No one of these models is totally convincing, but they make it possible to write down coherent, rigorous, and often elegant models of economies subject to increasing returns. So increasing returns are no longer something to be avoided or assumed away at all costs. The new intellectual opportunities offered by this revolution in theory have in turn transformed a series of other fields. In international economics the past decade has seen a virtually complete rethinking, with the emergence of a new view in which much trade represents arbitrary specialization based on increasing returns, rather than an effort to take advantage of exogenous differences in resources or productivity.[3] More recently, growth theorists have reintroduced the idea that sustained growth may arise from the presence of increasing returns, and old concepts like the "big push' have regained intellectual respectability.[4] And very recently some macroeconomists have suggested that increasing returns play a crucial role in business cycles.[5]

I believe that the time has come to use the same new tools to resurrect economic geography as a major field within economics. It is no longer the case that the need to model increasing returns makes a field untouchable. Instead, increasing returns are, for the moment at least, actually fashionable. And

---

3. See Helpman and Krugman 1985 for a survey of most of the concepts of the "new international economics."

4. See in particular Romer 1985,1987,1990 and Murphy, Schleifer, and Vishny 1989a.

5. See Hall 1989 and also Murphy, Schleifer, and Vishny 1989b.

so we can now admit to ourselves that space matters and try to bring geography back into economic analysis.

There are three reasons in particular why it is important to start doing economic geography. First, the location of economic activity within countries is an important subject in its own right. Certainly for a large country like the United States, the allocation of production between regions is an issue as important as international trade—a d more important than many issues that occupy a much larger part of economists' time. (I have my favorite candidates, but I won't tell you what they are; I have to live with these people for the next thirty years.)

Second, the lines between international economics and regional economics are becoming blurred in some important cases. One need only mention 1992 in Europe: as Europe becomes a unified market, with free movement of capital and labor, it will make less and less sense to think of the relations between its component nations in terms of the standard paradigm of international trade. Instead the issues will be those of regional economics—and it will help if we actually have some interesting regional economics to offer when the time comes.

To my mind, however, the most important reason to look again at economic geography is the intellectual and empirical laboratory that it provides. The "new" trade, growth, and business cycle theories of the past decade have suggested to us a world view of economics that is very different from that of

most pre-1980 theory. Pervasive increasing returns and imperfect competition; multiple equilibria everywhere; an often decisive role for history, accident, and perhaps sheer self-fulfilling prophecy: these are the kind of ideas that are now becoming popular. Yet it is very difficult to produce compelling evidence from trade, growth, and business cycles that this is the way the world really works. I at least am convinced that there is a strong arbitrary, accidental component to international specialization; but not everyone agrees, and the limitations of the data make a derisive test difficult. Paul Romer is convinced that increasing returns play a large role in explaining sustained growth; but not everyone agrees, and even I am agnostic. Robert Hall thinks that increasing returns play a crucial role in business cycles (he argues that a city and a boom are essentially the same thing—one in space, one in time); not everyone agrees, and I for one find this totally implausible (but interesting!).

But when one turns to the location of production within countries, the evidence for what Nicholas Kaldor called "the irrelevance of equilibrium economics" is far more compelling. The long shadow cast by history and accident over the location of production is apparent at all scales, from the smallest to the largest—from the concentration of most U.S. manufacture of wind musical instruments in the tiny town of Elkhart, Indiana, to the fact that a third of the U.S. population still lives within the original thirteen colonies. And this dear dependence on history is the most convincing evidence available that we live

in an economy closer to Kaldor's vision of a dynamic world driven by cumulative processes than to the standard constant-returns model.

What I want to do in this lecture is to offer a first illustration of the importance of economic geography, both as a field in its own right and as a way to see what kind of economy we live in. In particular, I want to show two things: that increasing returns are in fact a pervasive influence on the economy, and that these increasing returns give a decisive role to history in determining the geography of real economies.

I have already suggested that increasing returns affect economic geography at many scales. At the bottom of the scale, the location of particular industries—autos in Detroit, chips in Silicon Valley—clearly often reflects the "locking in" of transitory advantages. At an intermediate level, the existence of cities themselves is evidently an increasing returns phenomenon. At the grand level, the uneven development of whole regions (which in the United States may well be bigger than European nations) can be driven by cumulative processes that have increasing returns at their root.

In this lecture series I will pass over the question of urbanization relatively lightly. It has been better studied than the other issues I will consider (urban economics is more of an accepted field than economic geography), and it is also less relevant than the other aspects to international trade, which remains my ultimate interest. So I will focus on the small and the large:

the localization of particular industries and the differential development of huge regions. Today we look at the large, next lecture on the small.

To introduce the subject of divergent regional development, I turn to economic history to provide a particularly clear-cut example of the forces of economic geography at work. I then offer a simple model that helps make sense of that example. The example is the case of the U.S. "manufacturing belt": a relatively narrow stretch of territory within which the preponderance of U.S. manufacturing was concentrated from the mid-nineteenth century until the 1960s. The model—which is here developed only sketchily—is one in which the interaction of demand, increasing returns, and transportation costs drives a cumulative process of regional divergence.

The Case of the U.S. Manufacturing Belt

Early in this century, geographers noted that the great bulk of U.S. manufacturing was concentrated in a relatively small part of the Northeast and the eastern part of the Midwest—roughly speaking, within the approximate parallelogram Green Bay—St. Louis—Baltimore—Portland (figure 1.1). This "manufacturing belt"[6] took shape in the second half of the nineteenth

[6.] The term was apparently fast used by DeGeer (1927). The belt is not unique, nor are the forces that established it confined to national boundaries. Industrial Canada, concentrated in part of Ontario, is essentially a part of the U.S. manufacturing belt. Continental Europe has a 'manufacturing triangle' containing; the Ruhr, Northern France, and Belgium that is a close cousin of the U.S. belt.

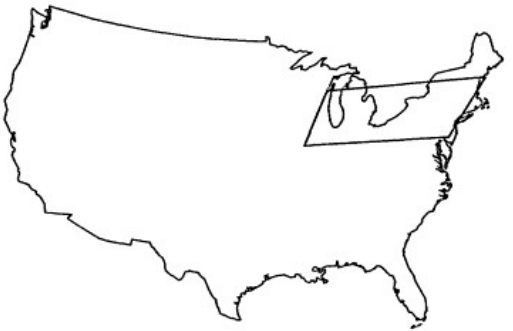

Figure 1.1

century and proved remarkably persistent. Perloffetal. (1960) estimated that as late as 1957 the manufacturing belt still contained 64 percent of U.5. manufacturing employment—only slightly reduced from its 74 percent share at the turn of the cent.

Even this number understates the manufacturing dominance of this region, because during the heyday of the belt most of the manufacturing outside it consisted either of processing of primary products or of production for a very local market. That is, the manufacturing belt contained virtually all manufacturing that was "footloose," not tied to other locations either by the need to be very close to the consumer or by the need to use natural resources very close to their source.

Why did the manufacturing belt play such a dominant role for so long? It was clearly not a case of an enduring advantage in natural resources: the manufacturing belt persisted even as the center of gravity of agricultural and mineral production shifted far to the west. In 1870 the Northeast and East North Central regions—within which the emerging manufacturing belt lay—accounted for 44 percent of U.S. "resource extraction" employment (agriculture, mining, forestry, fisheries). By 1910 this share had already fallen to 27 percent; yet these regions still accounted for 70 percent of manufacturing employment. And whereas the belt's share of manufacturing employment understates its manufacturing dominance, its share of resource employment overstates its resource base. The reason is that much of the agriculture that took place within or near to the manufacturing belt was quite different from that outside it: it consisted largely of truck farming and dairying, existing less because of the suitability of the land than because of proximity to the urban centers. In other words, if the manufacturing belt had not existed, the Northeast and Great Lakes areas would have had an even smaller share of agricultural employment.

H. H. McCarty, writing during the belt's heyday, summarized the divergence between regions bluntly: "Outside the manufacturing belt, cities exist to serve the farms; inside, farms exist to serve the cities."

As for mineral resources, the manufacturing belt originally drew some of its critical raw materials from nearby coal mines

and oil wells. By the mid-twentieth century, however, the great bulk of the raw materials for manufactures were imported from other regions.

Why, then, did so much of U.S. manufacturing stay within this relatively small stretch of territory? The answer in broad terms is, of course, obvious: each individual manufacturing facility stayed within the manufacturing belt because of the advantages of being near other manufacturers. And the apparent incentive for manufacturers to duster together explains the persistence of the manufacturing belt even after the bulk of U.S. primary production had shifted to other regions. Once the belt had been established, it was not in the interest of any individual producer to move out of it.

One may ask why this geographical concentration became established in the first place—a question about historical specifics to which I will return below. First, however, let us ask the more fundamental question: what were the forces that led manufacturers to want to duster together? I will sketch out a simple model in which geographical concentration results from demand externalities. This surely does not capture the full story, but it is strongly suggestive of the kind of explanation that is needed.

A Model of Geographic Concentration

The basic story of geographic concentration that I will propose here relies on the interaction of increasing returns, transporta-

tion costs, and demand.[7] Given sufficiently strong economies of scale, each manufacturer wants to serve the national market from a single location. To minimize transportation costs, she chooses a location with large local demand. But local demand will be large precisely where the majority of manufacturers choose to locate. Thus there is a circularity that tends to keep a manufacturing belt in existence once it is established.[8]

Imagine a country in which there are only two possible locations of production, East and West, and two kinds of production. Agricultural goods are produced using a location specific factor (land), and as a result the agricultural population is exogenously divided between the locations; for the moment we assume that the division is fifty-fifty.

Manufactured goods (of which there are many symmetric varieties) can be produced in either or both locations. If a given manufactured good is produced in only one location, transportation costs must be incurred to service the other market. On the other hand, if the good is to be produced in both

7. This lecture presents only a sketch of a model. It will be apparent that this sketch is sloppy about a number of issues, including, What is the market structure in manufacturing? What happens to profits, if any? and What resources are used in both fixed costs and transportation? It is possible to derive similar results in a fully specified general equilibrium monopolistic competition model; such a model is presented in appendix A. I adopt the more ad hoc approach here for ease of exposition.

8. In this model I stress the role of demand in determining the location of production of goods that are traded interregionally. An alternative approach would stress the role of increasing returns in the production of nontraded goods, as in Faini 1984. Eventually deciding between approaches will have to be an empirical matter; but for now it is a matter of taste.

locations, an additional fixed setup cost is incurred. The manufacturing labor force in each location is proportional to manufacturing production in that location. Finally, assume that the demand for each manufactured good in each location is strictly proportional to that location's population.

The basic idea can then be illustrated with a simple numerical example. Suppose that 60 percent of a country's labor force are farmers, divided equally between East and West. Suppose also that the total demand for a typical manufactured good is 10 units. Then if all manufacturing is concentrated in one location, that location will demand 7 units (3 demanded by the local farmers, 4 by the manufacturing workers), while the other demands 3; if manufacturing is evenly divided between the locations, each location will offer a local demand of 5.

To figure out what happens, we need to specify the fixed costs and transportation costs; suppose that the fixed cost of opening a plant is 4, and that the transportation cost per unit is 1. Then we have the situation shown in table 1.1. The table shows the costs to a typical firm of three locational strategies, contingent on the locational strategies of all other firms. Thus suppose that all other manufacturing is concentrated in East. Then our firm will have a local demand in East of 7 units, a local demand in West of only 3 units. If it serves the national market from a single plant in East, it will incur a fixed cost of 4 and a transport cost of 3. This is obviously less than serving the national market from a plant in West, which will have the same

Table 1.1
A manufacturing location story

Distribution of manufacturing employment	Costs of typical firm if it produces in			
		East	Both	West
East only	Fixed	4	8	4
	Transportation	3	0	7
	Total	7	8	11
Fifty-fifty split	Fixed	4	8	4
	Transportation	5	0	5
	Total	9	8	9
West only	Fixed	4	8	4
	Transportation	7	0	3
	Total	11	8	7

fixed cost plus a transport cost of 7; it is also less than building a plant to serve each local market, which saves the transport cost but incurs a double fixed cost of 8. In this case, then, the typical firm will choose to produce in East for a national market.

If each firm concentrates its production in East, however, then manufacturing production as a whole will be concentrated in East—which is what is assumed. So concentration of production in East is an equilibrium.

But it is not the only equilibrium. As the rest of the table shows, if manufacturing is concentrated in West, each firm will similarly also want to concentrate its production in West. And if production is split between East and West, each firm will want

to split its production, too. So in fact all three distributions of production—all in East, all in West, and a fifty-fifty split—are equilibria in this example.

The possibility of multiple equilibria can also be seen graphically (figure 1.2). On the horizontal axis we measure the share of the manufacturing labor force employed in West, on the vertical axis the share of West in the total population. The line *MM* represents the dependence of the distribution of manufacturing on the distribution of population; the line *PP* the converse effect of manufacturing on population distribution.

Let's begin with *PP*. This line represents the relationship between manufacturing labor force employment and total population. Let be the share of the total population engaged in manufacturing, let s_M be the share of the manufacturing labor force employed in West, and let s_N be West's share of the total population. West is home to half of the farmers, so that at

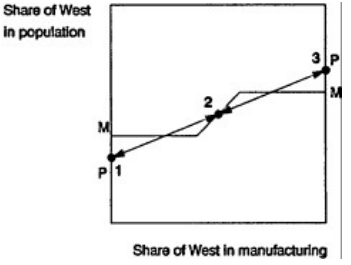

Figure 1.2

minimum it has a population share of (1-)/2. The more manufacturing it has, the larger this share:

$$s_N = \frac{1-\pi}{2} + \pi s_M.$$

This is an upward-sloping line that is, however, flatter than a 45-degree line.

Next turn to *MM*. Suppose that West has a very small share of the population. Then it will not be worthwhile incurring the fixed costs of establishing a manufacturing facility there; it is cheaper to serve the market from facilities in East. Conversely, if West has a very large share of the population, it is not worth producing manufactures in East. If the fixed cost is not too large relative to transportation costs, a sufficiently equal division of population will lead manufacturers to produce locally for both markets. Putting these observations together, we get the illustrated shape of *MM*: no Western production for low Western population, production proportional to population for intermediate levels, no Eastern production if the West is big enough. Let x be the sales of a typical manufacturing firm, F the fixed cost of opening a branch plant, and t the transportation cost of shipping a unit of manufactures from East to West or vice versa. Then it is cheaper to service West from a plant in East than to open a Western plant as long as $s_N xt << F$; it is cheaper to service East from West if $(1-s_N)xt << F$; and it is cheaper to have a plant in each region if neither is true.

Provided that fixed costs are not too high relative to transport costs,[9] we therefore have

$$s_M = 0 \text{ if } s_N < \frac{F}{tx}$$
$$= s_N \text{ if } \frac{F}{tx} < s_N < 1 - \frac{F}{tx}$$
$$= 1 \text{ if } 1 - \frac{F}{tx} < s_N.$$

Suppose that manufacturing production adjusts gradually toward its equilibrium level. Then the dynamics are illustrated by the arrows in figure 1.2. There are three stable equilibria: manufacturing may be concentrated in either location, at 1 or 3, or it may be equally divided, at 2. Which equilibrium you get to depends on where you start: history matters.

Of course there need not be multiple equilibria. The concentration of production, if it happens, depends on a demand externality. Manufacturers want to locate where the market is largest; the market is largest where the manufacturers locate. This circularity, however, need not always be strong enough to prevail over the pull of the dispersed agricultural sector. The situation could instead look like figure 1.3: a unique, stable equilibrium with manufacturing equally divided between the two locations.

9. If $F > tx/2$, then it is always cheaper to service both markets from a single plant, even if the population is equally divided. In this case *MM* is simply a horizontal line, and the possibility of an equilibrium with equally divided manufacturing disappears.

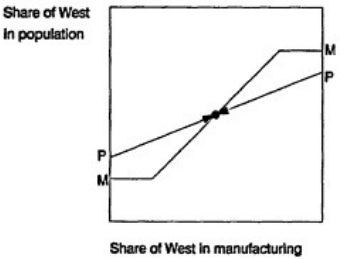

Figure 1.3

We can easily derive a necessary condition for concentration of manufacturing production in one location. With all manufacturing in East, West has a share of total population equal to only (1-)/2. The transportation cost of serving this market from East for a typical manufacturer is therefore $t x(1-)/2$. The cost of setting up a plant in West is F. So a concentration of production in East, once established, will persist as long as

$$F > \frac{1-\pi}{2} tx .$$

If this criterion is not met, history does not matter: the geography of manufactures will follow that of agriculture.

We can immediately see that a key role for history depends on three parameters: large F, i.e., sufficiently strong economies of scale; small t, i.e., sufficiently low costs of transportation; and

large , i.e., a sufficiently large share of "footloose" production not tied down by natural resources.

We can now tell a stylized story of the emergence of the manufacturing belt.[10] In the early United States, with its primarily agricultural population, where manufacturing was marked by few scale economies and where transportation was costly, no strong geographical concentration could occur. As the country began its industrial transition, manufacturing arose in areas that contained most of the agricultural population outside the South—and the South was, for reasons having to do with its uniquely awful institutions, unsuited for manufacturing. During the second half of the nineteenth century, however, manufacturing economies of scale increased,[11] transportation costs fell, and the share of the population in nonagricultural occupations rose. The result was that the initial advantage of the manufacturing belt was locked in. Even though new land and new resources were exploited to the west, even though slavery ended, for three-quarters of a

[10] This story is based on the fascinating work of David Myers (1983), who however bears no responsibility for the crudity of the representation.

[11] Chandler (1990) provides a fascinationg story of the emergence of large manufacturing firms in the period between the Civil War and the 1920s—that is, during the heyday of the manufacturing belt. He shows that in one industry after another, a "first-mover" led the way by taking advantage of new technology and lower transport costs to build one or two plants of unprecedented size, serving the whole national market. While Chandler does not emphasize the point, his U.S. firms invariably established their first huge plant somewhere inside the manufacturing belt. Sometimes this choice was dictated by the availability of specific resources—for example, hydroelectric power for aluminum smelters at Niagara Falls—but access to markets seems to have played a key role in ruling out sites outside the manufacturing belt.

century the pull of the established manufactured areas was strong enough to keep the manufacturing core virtually intact.

Of course this story oversimplifies in a number of ways. On one side, it probably underemphasizes the role of certain conventional factors in giving rise to the manufacturing belt—there is a suspicious correlation between the location of heavy industry and that of coalfields, both in the United States and in Europe. On the other side, it says nothing about the sources of local specialization within the manufacturing belt—about why Detroit emerged as the automotive center, New York as the garment center, Grand Rapids as the furniture center, etc. Yet it surely captures an important aspect of what happened. And it also contains elements—increasing returns at the level of individual firms, and external economies resulting from the interaction of these firms' decisions—that will reappear as one further elaborates the story.

Before changing the subject, however, there is one particular aspect of the rise of the manufacturing belt that deserves some further elaboration. This is the role of the endogeneity of transport costs themselves.

Transport Networks and Regional Divergence

It is obvious from even a cursory reading of U.S. economic history that part of the advantage of the manufacturing belt arose from the density of the railroad network connecting the region's cities, a density that was itself a product of the region's

manufacturing dominance. This transport network effect deserves a little more attention.

Imagine for a moment a nation with not two but three locations—Center, West, and South—with equal transportation costs between any two locations. Where will a manufacturing firm locate? By analogy with our previous discussion, if one of these locations offers a sufficiently larger local market than the others, and if fixed costs are large enough relative to transport costs, the more populated location will attract a concentration of manufacturing production.

Now imagine a nation with *four* locations: East, Midwest, West, and South. But now suppose that the transportation cost between East and Midwest is much lower than that in other directions. Then in economic terms East and Midwest will in effect form a *single* location. The East-Midwest region will be a more attractive place to locate manufacturing than South or West, even if the individual markets are no bigger, because factories in either place will have better access to the combined market.

But why should transportation costs in one direction be much lower than in others? The most natural answer is that there are economies of scale in transportation itself. A railway or a highway represents indivisible investments, while the frequency of air service and the ability to use large, efficient planes depends on the volume of demand. Suppose that manufacturing production, and hence both demand and sup-

ply, is concentrated in East and Midwest. Then there will be a greater volume of transportation between these locations than on other routes. This will mean lower transport costs, which will in turn reinforce the advantage of East and Midwest as locations for production.

It is possible in principle to imagine this transportation network effect as an independent source of geographical concentration of industry—that is, to set up a model in which the local market size effect that is the driving force for our basic model is absent. In practice, of course, the two effects work together. The U.S. manufacturing belt was characterized not only by a denser population but also by a better transport network than any other part of the country, and thus offered much better market access to manufacturers.

Further Thoughts

The case of the U.S. manufacturing belt is of substantial interest in its own right. The rise and persistence of that belt is an important yet much neglected aspect of U.S. economic history. More important than its immediate significance, however, is what the history of manufacturing location says about the nature of our economy in general. And what it says is that increasing returns and cumulative processes are pervasive and give an often decisive role to historical accident.

It is also interesting that the story of the manufacturing belt reaches back to the mid-nineteenth century. It is common to

argue, as Brian Arthur has, that external economies and cumulative processes have become more important in recent decades because of the growing importance of technology. The geographical concentration of manufacturing in the United States took shape, however, long before the dawn of the information age. So it is not simply true that our economy is not now well described by the conventional constant-returns model It never was.

The Process of Change

The circular relationship in which the location of demand determines the location of production, and vice versa, can be a deeply conservative force, tending to lock into place any established center-periphery pattern. In the case of the U.S. manufacturing belt, the geographical structure of production that happened to exist at the point at which industrialization, factory production, and the railroad came into force remained essentially intact for the next century.

Nothing, however, is forever. Indeed, one of the most interesting things about the type of model sketched in this lecture is what is says about the process of economic. change. What I want to do at this point is illustrate two ideas in particular that are suggested by the center-periphery model. First is that while the geographical structure of production may be stable for long periods of time, when it does change it may change rapidly. In fact, a gradual change in underlying conditions can

at times lead to explosive, or more accurately, catastrophic change. Second, change when it comes may be influenced strongly not only by objective conditions but also by expectations—expectations that may be self-fulfilling.

The Logic of Sudden Change

To see how change in the geography of production can sometimes take place abruptly, suppose that instead of being divided equally between the locations, the agricultural labor force is unevenly split, with West initially having a smaller population. The hypothetical position is shown in figure 1.4. *PP* is the initial relationship between manufacturing employment and total population. Although a possible equilibrium exists at 2 in which West would produce manufactures, we suppose that owing to a head start for East we are instead at 1, with no manufacturing production in West.

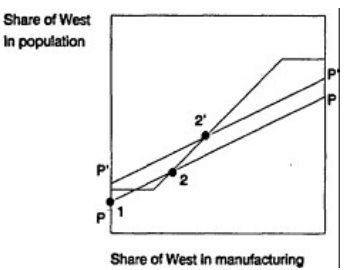

Figure 1.4

Now suppose that there is a gradual reallocation of the agricultural labor force from East to West. This will shift *PP* upward, toward *P'P'*. It is immediately obvious that at a certain point East's dominance of manufactures will collapse. When the Western population reaches a critical mass, it becomes worthwhile for manufacturers to produce there; as manufacturing production in West increases, the population rises further, stimulating still further increases in manufacturing production. A small increase in the agricultural base may therefore set in motion a cumulative process of import substitution and growth, leading eventually to an equilibrium at a point like 2'.

This scenario may not be entirely hypothetical. Paul Rhode (1988) has pointed out that late nineteenth-century California was a resource-based economy with limited manufacturing, largely because the local market was too small to support much industry. He suggests that the discovery of oil around the turn of the century raised California to critical mass, starting it on a process of explosive growth (and in particular causing the rapid emergence of Los Angeles as a manufacturing center).

The point is that the same kind of model that helps explain why history matters also suggests that when change does come, it will often be sudden. And we may also note that changes in regional fortunes will be difficult to predict: in the hypothetical history illustrated in figure 1.4 one would see a

sudden acceleration in West's growth without any obvious reason.

History versus Expectations

Now that I have described the logic of sudden change, let me raise a problem with that logic—a problem that will no doubt already have occurred to those readers with a background in modem macroeconomics, with its emphasis on rationality of expectations. Suppose that the distribution of agricultural population were in fact evolving in the way illustrated in figure 1.4. Wouldn't manufacturing workers and/or firms realize that a sudden increase in West's population was in prospect? And wouldn't they begin to move into West in anticipation of that increase, thereby smoothing out the process of change?

The answer is yes, if they were sufficiently well informed. In practice I have my doubts—here as elsewhere, the assumption of rational expectations seems to presume a degree of information and sophistication that is unreasonable. That is not a controversy that I want to get too far into; I just want to argue that the kind of static expectations that implicitly underlies the dynamics in figure 1.4 retains a useful place in analysis.

But even if I am skeptical about the literal relevance of the assumption of rational expectations, now that we have raised the issue of the role of expectations in regional development,

we should pursue it. For if you think about it a bit, you realize that the kind of circular process that I have argued leads to regional differentiation can also lead to self-fulfilling prophecies.

Imagine again our two-region nation; this time assume for simplicity that the economies of scale are sufficiently large relative to transport costs that there are only two long-run equilibria, with manufacturing concentrated either entirely in East or entirely in West. Suppose, however, that workers cannot all move at once; that there is some kind of adjustment cost that limits the rate at which manufacturing can shift. Thus a worker who chooses to locate in one region or the other is stuck with that choice, at least for a while.

It is immediately apparent that in this case workers will be concerned with more than their current wage—they will base their decisions to move on something like the present value of future wages. But the real wage rates in each region at any point in time depend on the distribution of manufacturing workers; so this means that each worker's current location decision depends on her expectations about the future decisions of other workers.

The possibility of self-fulfilling prophecy now becomes apparent. Suppose that East and West have equal numbers of farmers, and that East initially has somewhat more manufacturing, so that by virtue of its superior forward and backward

linkages East offers a higher real manufacturing wage. Then one might expect to see migration of manufacturing from West to East. But suppose that for some reason the public is convinced that West will be the destination of migration, not East, and that as a result real wages in West will eventually exceed those in East. This belief will induce seemingly perverse migration from the region with higher real wages to that with lower—and this migration will indeed eventually reverse the real wage differential! And if this reversal takes place sufficiently quickly, the worker who migrates from East to West will find that she actually made the right decision. Thus the belief that West is the land of opportunity turns out to be a self-fulfilling prophecy. If everyone had instead had faith in the East, of course, East would have gotten the industry.

When can self-fulfilling prophecy outweigh initial advantage? Several factors clearly matter. First, the rate at which workers and firms can move must be rapid enough relative to the rate at which future wage differentials are discounted that the future advantage of one region can matter more than the current advantage of another. Second, increasing returns must be strong enough that an expected future shift in population distribution moves the real wage differential quickly. Finally, the starting position must not be too unequal: if enough manufacturing is concentrated in one region, this initial advantage may be too much for even the most optimistic expectations about the other region to overcome.

It is possible to formalize the problem of self-fulfilling expectations rather neatly; such a formalization is presented in appendix B. What the formalization tells us is that there may be a range of initial distributions of manufacturing workers from which either region can end up with the manufacturing concentration, depending on expectations. Whether such a range exists, and the size of the range if it does exist, depend crucially on the speed of adjustment; only if the adjustment is slow can we be sure that initial advantage cumulates over time, instead of potentially being overruled by self-fulfilling expectations.

So much for the logic. To what, if anything, does this story correspond in reality? The answer is that I am not sure. In the case of the U.S. manufacturing belt, history dearly determined what happened. Perhaps there could have been a self-fulfilling belief in the industrial future of, say, the Great Plains that could have outweighed the historical advantage of the traditional manufacturing locations; but there wasn't. (And I doubt it.)

At a smaller scale, however, the case for self-fulfillment is better. Certainly a prominent part of the tradition of local economic development in the United States has been boosterism—the sometimes ludicrous efforts by local businessmen and chambers of commerce to convince footloose individuals and firms of the virtues of their state or town, in the belief that if they can draw a critical mass into the local economy, it will become self-sustaining. Some of this booster-

ism involved concrete incentives, sort of proto-industrial policy; we'll see an example in the story of Akron and the rubber industry in the next lecture. But often it was simply an attempt to create optimism about the locale. The analysis sketched out here suggests that in principle, at least, boosterism may make perfectly good sense.

There is also the possibility of reverse boosterism: if for some reasons businesses and workers become pessimistic about a region's prospects, this pessimism can become self-justifying. It is difficult for me to avoid speculating that something like this may be happening in my own home state. As you may know, two years ago the governor of Massachusetts ran for president, in part on the impressive economic record of his state. He was humiliatingly defeated by George Bush—and the Massachusetts economy itself went into a tailspin. Was this just a coincidence, or did the psychological impact of the campaign, and the political civil wars that followed within the state, create a self-fulfilling downward spiral? (And will the state's economy continue to implode?) I don't know the answer, but such seemingly fanciful ideas don't seem as silly to me as they might to a more conventionally minded economist.

Where We Stand

In this lecture I have tried to argue for the acceptance of economic geography as a major field within economics, on a par with or even in some sense encompassing the field of

international trade. Since we all know that economic argumentation succeeds at least as much on its aesthetics as its empirical support, I have tried to make my case with the cutest model of geography that I have been able to come up with: one that shows how a core-periphery structure can emerge endogenously on a nationwide scale. And I have argued that something like this actually happened in the United States between the Civil War and the First World War.

The phenomenon of concentration in economic geography takes place at many scales, however. While the emergence of huge metropolitan belts may be the most dramatic, for international affairs the forces that lead to localization of particular industries, usually but not always within those belts, are possibly of even more interest. So in the next lecture I will move from the large to the small: from core-periphery to localization.