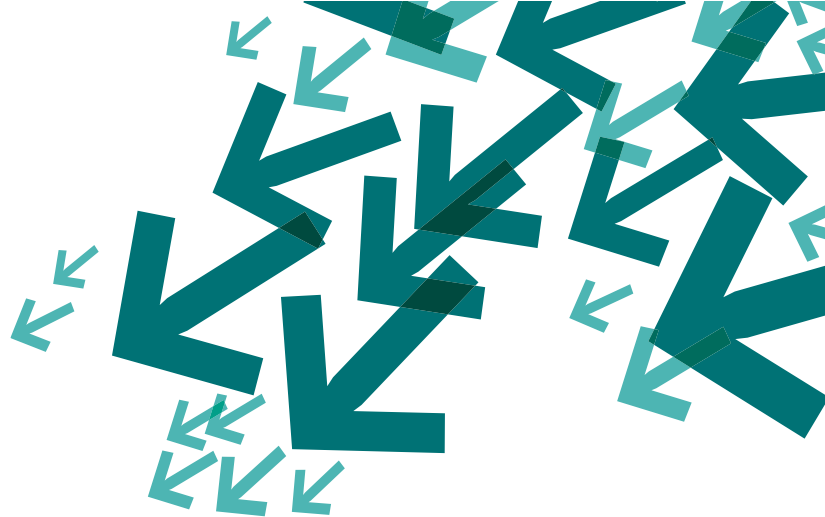




# Shifts in economic geography and their causes

**R**ecent decades have seen momentous changes in the economic geography of the world. Political transitions and economic liberalisation have brought formerly closed countries into the world economy. In Richard Freeman's phrase, this amounts to a 'doubling of the world labour force': the collapse of Soviet communism adds 260 million workers, the opening up of China 760 million, and Indian liberalisation 440 million.

At the same time, technological change has continued to reduce the cost of interactions within and between countries. Accompanying technological innovation has been business innovation. Multinational firms have expanded rapidly, with foreign direct investment growing at twice the rate of world trade, which itself has grown at twice the rate of world income. New forms of trade have emerged with the growth of outsourcing and production networks.

There have also been changes in the

location of economic activity. Figure 1 illustrates the historical record of shares of world GDP for different regions. In essence, there have been four phases: the initial dominance of Asia; followed by the rapid growth of Europe during and after the industrial revolution; then the subsequent rise of North America; and now the resurgence of Asia.

Part of the change is due to population, but much the larger part is due to changes in per capita income – the 'great divergence', which saw the ratio of per capita incomes of the richest to poorest nations increase from 8:1 in 1870 to more than 50:1 in 2000.

Economic geography has changed at all spatial scales – not just aggregate regions, but also within regions and countries. The most important sub-national change is urbanisation with a majority of people now living in cities. China alone expects a doubling of its urban population to nearly one billion people by 2030.

This sketch describes some of the forces driving change in the world

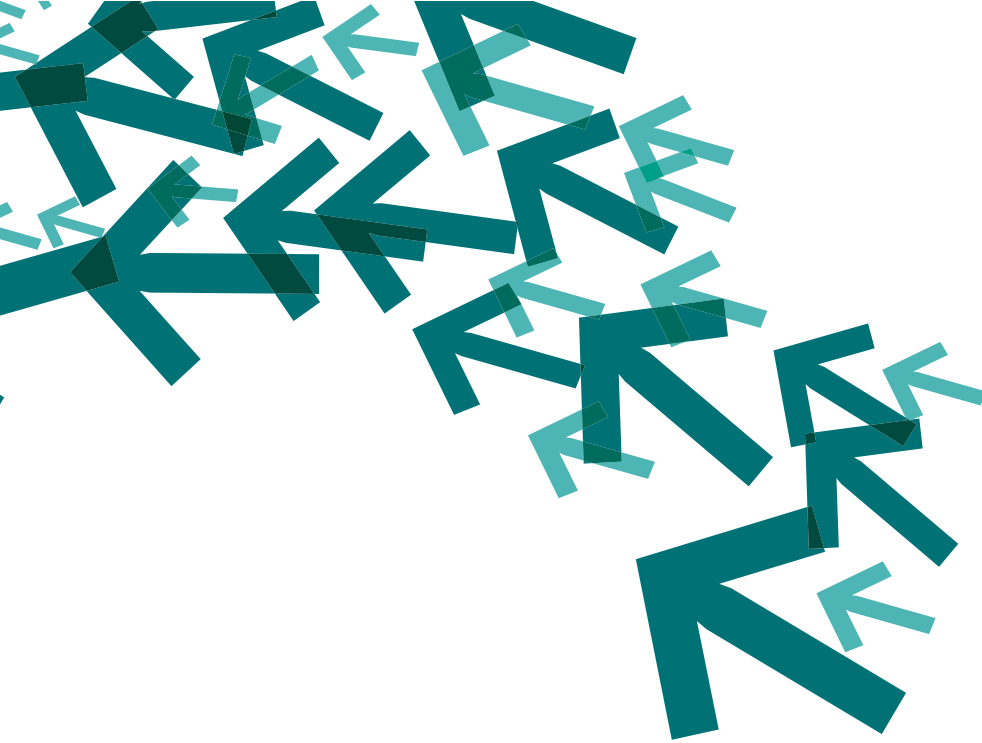
economy, and some of the ensuing changes in economic geography. It also challenges our understanding of the location of economic activity and the determinants of changes in the pattern of location.

## The key questions

The first question is: why are economic activity and prosperity spread so unevenly? Is an American really 50 or 100 times more productive than an Ethiopian? Even within the UK, why are a Londoner's earnings 70% higher than those of someone from Stoke?

Standard economic theory suggests that while differences may arise as some countries or regions gain initial advantage, they should be rapidly arbitrated away. Capital will flow to where labour is cheap, and knowledge and new technologies will be transferred. Fundamentally, if there are diminishing returns to economic activities, then there will be a continuing process of convergence and a tendency for activity to be spread relatively uniformly across space.

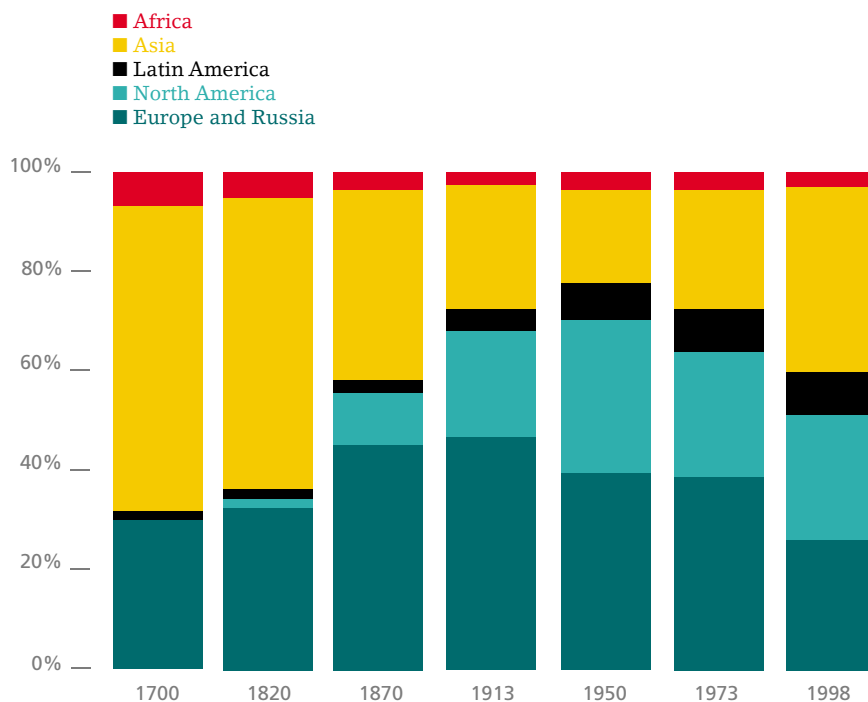
Yet these forces seem to operate in a



Why are economic activity and prosperity spread so unevenly, and does globalisation necessarily narrow these differences?

**Tony Venables** outlines the key forces driving the economic geography of cities, nations and the world economy – and how we should think about future patterns of location for both developed and developing regions.

Figure 1:  
Shares of GDP



Source: Angus Maddison (2001)

manner that is, at best, selective. Even within reasonably well-functioning market economies spatial disparities persist, and in the wider world economy they are astonishingly large.

The second question is: does increasing trade necessarily narrow these differences? The standard answer of international economics is that price and income differentials will narrow although production structures may diverge. Yet there have been many episodes in history where falling transport costs and opening to trade have been associated with increasing differences between countries.

So the third question is: how should we think about future developments, both for developed and for developing regions? We are currently in an era of globalisation in which some economic activity is dispersing from existing centres, but what determines which sectors move and where they go? Will prosperity be widely dispersed or will some regions continue to be left behind?

Answering these questions requires a general theory of the location of economic activity. This is not an area well served by mainstream economic theory, which is generally 'aspatial': the core theory of international trade treats countries as points with no spatial structure either between or within them. In addition, standard theory generally predicts convergence of performance: differences between regions arise because of 'exogenous' factors – differences in endowments, institutions or technology – and economic forces generally erode them or at least do not amplify them.

### New economic geography

'New economic geography' has worked towards addressing these deficiencies. The objective is to offer an integrated theory of location, capable of explaining divergence as well as convergence in economic performance.

The key building block is the recognition that proximity is good for productivity: dense configurations of economic activity work better than sparse or fragmented ones. Mobile factors of production – firms and possibly workers – will locate in order to take advantage of higher productivity, and this creates a positive feedback. Firms and workers go where productivity is high, and by so doing

tend to raise productivity further, creating an uneven distribution of activity and spatial income disparities.

There are several analytical challenges here. One is to have sound theories and evidence as to why proximity is good for productivity. The second is to place the proximity-productivity relationship in a wider model and thereby identify the trade-offs between forces for concentration or dispersion of economic activity. Then hypotheses can be formed about circumstances in which activities may concentrate or disperse, and about the associated shifts in economic geography.

The next three sections discuss three key propositions. First, that proximity to other economic agents – workers, consumers and firms – is good for productivity. Second, that large income disparities are a perfectly natural outcome of a world in which proximity matters. And third, that the effects of increased trade are potentially ambiguous: there are circumstances in which cheaper spatial interactions cause inequality not convergence.

## Proximity and productivity

The first proposition is that proximity to other people is good for productivity. What is the evidence, and what are the economic mechanisms that drive the effect? There is a long list of mechanisms, which fall under two headings: product markets and labour markets. These correspond loosely to different spatial scales: some of the product market effects might operate over long distances; while labour market effects are short range, even coming down to the benefits of face-to-face contact.

### *Product markets*

The most immediate effect of proximity is that it saves transport and other trade costs. Thus, if two producers have identical physical productivity, the one producing in the large market will have higher productivity since it does not have to bear the costs of transport to remote consumers. This producer will also have lower cost intermediate inputs, not having to absorb its share of shipping costs on these inputs.

But trade costs should be thought of in much more general terms than just

freight charges. Time in transit is costly – partly from the costs of carrying stock, and also from the likelihood that long transit times reduce the reliability and predictability of deliveries. It also makes firms slower to respond to changing demand or cost levels, which by itself can be a force for the clustering of activities (see Harrigan and Venables, 2006).

Transport and trade cost savings are a direct benefit of proximity, but its full economic impact comes from economies of scale associated with operating in an area of dense economic activity – close to consumers, workers and other firms. In a small or fragmented market, there is a trade-off between having firms large enough to achieve economies of scale without becoming monopolists.

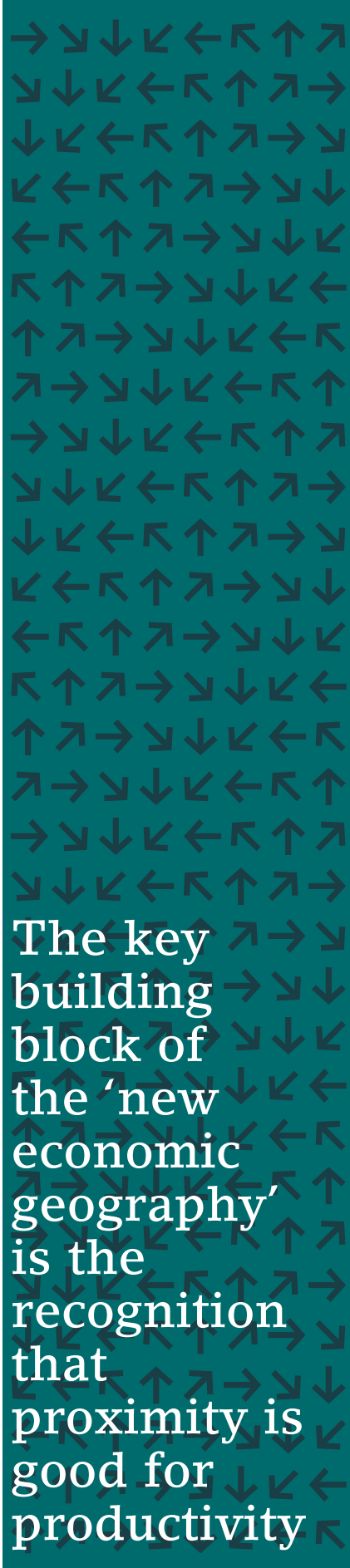
A large or integrated market shifts this trade-off, allowing benefits of both large scale and more intense competition. As a consequence, firms will be larger, operating at lower average cost and setting lower prices. The more intense competition will weed out less efficient firms, concentrating production in efficient firms. A larger market will also support a greater variety of products. These price and variety effects benefit consumers and, if the goods are intermediates, benefit firms in downstream sectors.

### *Labour markets*

In addition to efficiency gains in the goods market, firms also gain from operating in a large labour market. The larger the pool of workers that a firm can access, the more likely it is to be able to find the exact skills that suit its needs.

A large labour market will also increase the incentives for workers to undertake training. This argument turns on increased intensity of competition. In a small market, workers who acquire specialist skills may be 'held-up' by monopsonistic employers, so there is no incentive for them to invest in skills. A large number of potential employers removes this threat of opportunistic behaviour, and thereby increases training incentives.

A further set of arguments has to do with communication between workers. In many activities, face-to-face contact is extremely important, enabling higher frequency interchange of ideas than is possible by email, phone or videoconference. Brainstorming is hard to



The key building block of the 'new economic geography' is the recognition that proximity is good for productivity

do without the ability to interrupt and use parallel means of communication – oral, visual and body language.

Face-to-face contact is also important for building trust. By breaking down anonymity, it enables networks of the most productive workers to develop, and promotes partnerships and joint projects. All these considerations are productivity enhancing.

A final set of arguments concerns 'knowledge spillovers', which are easier between proximate firms than remote ones. The mechanism may be labour mobility, face-to-face social contact between workers or observation of the practices of other firms. Such effects are particularly important in innovation intensive activities.

Location specific knowledge spillovers also arise as firms learn about the characteristics of their location, and this knowledge spills over to other firms. This may be learning about real economic characteristics of locations, or may just be 'herding', as firms simply copy the location decisions of other (successful) firms. All of these knowledge spillover effects are summarised in Alfred Marshall's phrase: 'the mysteries of the trade become no mystery but are, as it were, in the air'.

### *Different scales of proximity effects*

The various proximity effects operate over quite different spatial scales. Product market effects can be long range: firms in New York may benefit from a large market in California, and reductions in international shipping costs will increase market access for exporting firms. Labour market effects operate within a much narrower area – indeed, Rice et al (2006) suggest that 45 minutes driving time is the appropriate range for these effects.

Proximity effects also operate across different sectoral scales. Some are driven by aggregate demand: proximity to a mass of consumers will cut trade costs and raise demand for all firms whose sales, direct or indirect, are concentrated in the area. All such firms will appear to have higher productivity near centres of high demand. Other effects are narrowly sector specific: for example, a film actor benefits from proximity to a film producer but won't care much about aggregate demand in Los Angeles.

The financial sector provides a good

example of these varying scales. Some of its backroom activities can be easily separated from the rest of the firm, and operated from low cost locations. Other parts of the business, such as retail banking, require proximity to final consumers. And the most skill intensive parts of the sector are spectacularly prone to clustering, valuing face-to-face contacts, access to thick labour markets and a dense network of firms offering complementary services.

Similarly, within manufacturing, some stages of the production process can be outsourced and moved to low cost locations. For other parts, this is not possible, partly because of the costs of breaking the production flow within the firm and partly because of the loss of proximity to complementary inputs, skilled labour markets or consumers.

### *The evidence for proximity effects*

There are many sources of evidence for the claim that proximity is good for productivity. The most extensively researched is from studies of the productivity of cities. Rosenthal and Strange (2004) report a consensus view that, over a wide range of city sizes, doubling city size is associated with a productivity increase of 3-8%. This is a large effect: moving from a city of 50,000 inhabitants to one of five million is predicted to increase productivity by more than 50%.

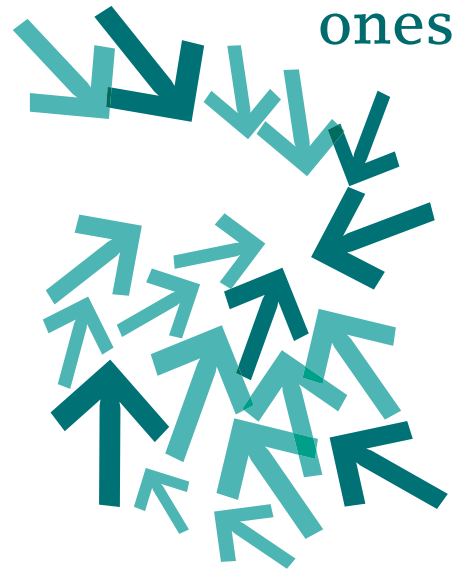
In the international context, proximity manifests itself in large trade flows – and the gains from trade are widely documented. For example, Redding and Venables (2004) focus on measuring countries' access to markets and sources of supply, and find that a 1% improvement in a country's market access – which has the effect of increasing its exports by 1% – raises per capita income by around 0.25%.

### **Equilibrium disparities**

The second proposition is that large spatial disparities in income can be a persistent 'equilibrium' outcome. To establish this, the arguments about proximity need to be combined with other forces to give a theory of the location of economic activity, and consequent wage and income differentials.

The best way to do this is to think

## Dense configurations of economic activity work better than sparse or fragmented ones



about the profitability of a firm choosing between various production sites. How do its potential profits vary across alternative locations? They depend on three elements: productivity, defined broadly to include the benefits of transport cost savings; product market competition – the number of competitors that the firm will face in its chosen location; and input prices, including those of intermediate goods and primary factors.

The equilibrium location of activity is the arrangement of firms that causes productivity levels, product market competition and input prices to adjust until all firms are indifferent about their choice of location.

Now what happens to the profits of firms in a location when an additional firm establishes operations in the same location? If profits increase, then adding this firm increases the incentives for further firms to come, so there is an agglomeration process, with differences between locations becoming amplified. If profits fall, then activity will be dispersed and firms will tend to spread out.

# Large income disparities are a perfectly natural outcome of a world in which proximity matters



The proximity-productivity relationship is an amplification force, since adding firms raises productivity and profits of existing firms. Product market competition and input prices are dispersion forces: adding another firm crowds the market, thus reducing revenue, and bids up the prices of immobile factors, raising costs.

Equilibrium location is therefore a balance between the proximity-productivity relationship, a force that amplifies initial differences, and product market competition and factor cost forces, which tend to dampen down effects.

## *First nature geography and international wage differences*

How do these forces interact? Consider first the implications of exogenous differences between countries, such as institutional or policy differences or differences in natural geography. Geographers have a longstanding distinction between 'first nature geography' – coasts, mountain ranges, natural endowments – and 'second nature geography' – the geography of

interactions between economic agents.

There are the direct disadvantages of bad first nature geography – for example, propensity to disease lowers productivity and being landlocked raises transport costs – but what are the full equilibrium effects? Advantages and disadvantages of first nature geography become amplified, as firms move into locations with good geography, and the proximity-productivity relationship causes further increases in productivity, while countries with bad first nature geography will have low levels of economic activity, reducing productivity further.

Who bears the costs and benefits of these spatial variations in productivity? They are borne entirely by immobile factors, which in the international context means labour. Since labour may be a small share of the costs of production, there can be a large multiplier effect. If labour is 10% of gross costs, then a 50% difference in the productivity of all inputs will translate into a 500% wage difference.

Such large effects are confirmed by Gallup and Sachs (1999), who find that 70% of cross-country variation in per capita income can be accounted for by just four measures of physical and economic geography: malaria, hydrocarbon endowment, coastal access and transport costs.

## *Second nature geography and economic agglomeration*

The proximity-productivity relationship does not just amplify economic differences that arise because of exogenous factors. If amplification effects are strong enough, then they can create disparities between locations that are identical in underlying characteristics. Indeed, they are the driving force behind the existence of cities, the most commonplace manifestation of the unevenness of economic activity.

A world with diminishing returns to activity would have no cities, as activity would be smeared across space. But the proximity-productivity relationship is a force for clustering all activity into a mega-city.

Pulling in the opposite direction are dispersion forces: product market competition, which means that some firms remain dispersed to supply remote consumers; and high urban prices of immobile factors. If workers are free to migrate within a country, then the only immobile factor is land, the price of which

is bid up, thus also raising urban wages as mobile workers are compensated for regional variations in the cost of living. Further dispersion forces may be provided by urban congestion and commuting costs.

Notice that dispersion forces are generally not sector specific but some agglomeration forces are since the proximity-productivity relationship can vary between sectors. This gives rise to sectorally specialised cities – London and Hollywood – the size of which depends on the importance of the sector in the world or regional economy.

## **Trade, location and inequality**

The third proposition is that trade is not necessarily a force for convergence of incomes. The historical record shows that nineteenth century globalisation was associated with substantial divergence of income between regions, and the impact of twenty-first century globalisation on international inequality remains hotly debated.

The interactions between trade and income divergence are complex, but the basic ideas can be developed using the stylised model of Krugman and Venables (1995). This model has just two countries, which have identical economic structures and identical real wages when trade costs are high. This is because when trade is expensive, supply and demand in each country's product market – a dispersion force – are dominant in determining the location of activity.

As trade costs fall, the possibility of supplying consumers through trade rather than local production develops, and the proximity-productivity relationship becomes relatively more important. Below some level of trade costs, these forces come to dominate, and one of the countries gains most of the manufacturing and the other is 'deindustrialised'.

But as trade costs fall further, so the clustering force becomes weaker, and location comes to be determined by factor prices, a dispersion force. This is the era of globalisation, in which manufacturing starts to move from developed to developing countries and wages narrow.

Clearly, this model is highly stylised, but it illustrates the complex role of trade in determining the location of activity. Trade changes the balance between the

dispersion forces of product and factor market competition and the clustering force of the proximity-productivity relationship. The model provides the apparatus to think through other shifts in economic geography.

### Lumpy dispersion

So world economic geography can be thought of as a balance between concentration forces and dispersion forces. What light does this shed on potential future changes in a globalising world economy?

This question can be answered under three headings. First, sectorally: which activities are likely to remain concentrated and which to disperse? Second, by country: what might the cross-country pattern of location look like, and how will the international distribution of income change? And third, sub-nationally: where next for cities?

The running theme is that much activity will move out of existing centres, but relocation will be 'lumpy', benefiting some regions more than others and re-coalescing into new patterns of agglomeration.

### Which sectors move?

Which sectors are most likely to detach from existing centres of activity and relocate to lower wage regions? One determinant is factor intensity: so, for example, unskilled labour intensive activities will tend to relocate to low wage countries. But it is helpful to extend this reasoning with a broader notion of comparative advantage, encompassing a variety of country characteristics, including institutional quality and business environment. Thus, countries with good intellectual property protection will tend to attract sectors that value this protection, and so on.

Comparative advantage is only part of the story. There is the 'linkage' intensity of the product: how easy is it to detach the activity from its existing location, and how expensive is loss of proximity to related economic activities? This depends on all the elements of the proximity-productivity relationship. If firms in a sector are highly dependent on a network of suppliers or on capabilities embodied in the local labour force, then it is unlikely that the sector will relocate.



The effects of increased trade are potentially ambiguous: cheaper spatial interactions can cause inequality not convergence

The strength of these interactions varies across sectors and depends on the costs of transport and other spatial interactions. Thus, face-to-face contact may be crucial for some economic activities but not for others. Skills may be embodied in the labour force and hard to transfer, or it may be very easy to train workers in a new location. And timely delivery may be crucial for some goods: there is evidence that production of fashion sensitive garments has moved back to high wage countries for this reason.

The profitability of relocation also depends on the extent to which the production process can be 'fragmented' into different stages, with different factor endowments and different linkages to related activities. This is being studied in a rapidly expanding research literature on fragmentation, production networks, outsourcing and offshoring (for example, Grossman and Rossi-Hansberg, 2006).

The argument is that globalisation has created the possibility of a finer pattern of specialisation as it is now possible to locate different parts of the production process in different countries. Component parts and semi-finished goods can cross borders multiple times, and countries are able to engage in 'vertical specialisation', producing just one very narrowly defined part of a product.

This is an area where much more research is needed. For high-income countries, is it possible to identify activities that are more or less likely to become detached? We need to develop a way of measuring whether a country has a 'deep' or a 'shallow' comparative advantage in a particular product or task.

For developing countries, which are the sectors that are most footloose – apparel, electronic assembly? And can we diagnose why a labour intensive activity might not be willing to move? What exactly are the linkages that would be forgone in moving out of an established centre of activity, and how easily can they grow in a developing country?

There are also a number of policy issues. The proximity-productivity relationship creates a coordination failure, suggesting a role for national industrial policy to act as a catalyst to overcome the failure. Internationally, it should influence the way we think about trade policy.

### *Where will production go?*

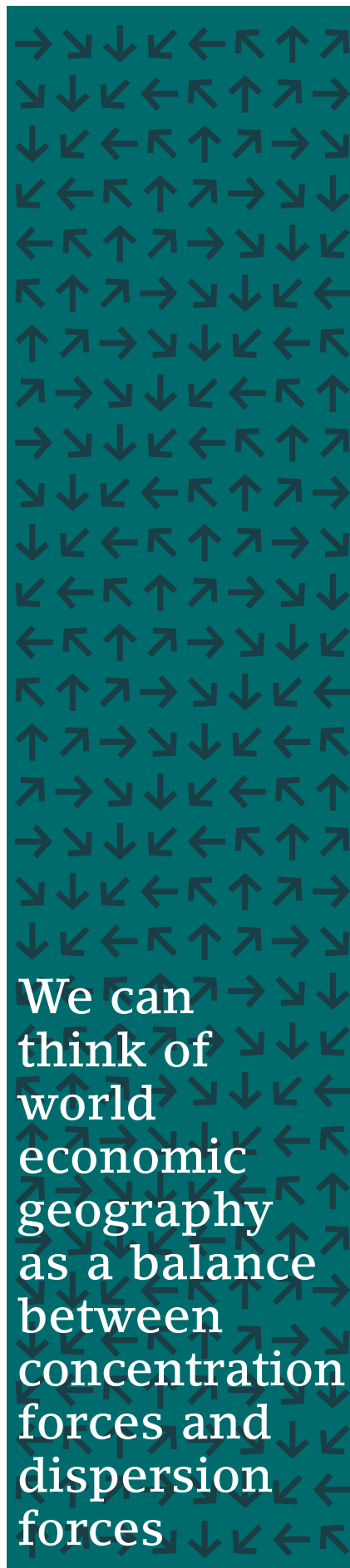
Turning to countries, what pattern of development is predicted? The fundamental point from the theory is that simultaneous development of similar countries is likely to be unstable. For example, suppose that activity is relocating from an established centre into two similar emerging economies, and that proximity-productivity relationships operate in the sectors concerned. Then whichever country gets slightly ahead will have higher productivity and become the more attractive location for further investment, while the other country will fall behind.

This observation has a number of implications. First, we should expect growth and development to occur in sequence, not in parallel. Instead of all poor countries steadily converging to high-income status, there is an inherent unevenness. Some countries will grow extremely fast while others will be left out of the process.

Which countries go first? Many factors count, including first nature geography and the institutional and policy environment. The models predict that economic development will spread out from existing centres, going to regions with low transport costs, such as the coastal regions of neighbouring countries. This is a view of the world that fits well with recent growth patterns in Asia as compared with Africa.

While this aggregate view is important, the phenomenon is seen even more sharply at the sectoral level. As sectors migrate from established centres of activity, so their new location pattern exhibits clustering. A striking feature of growth has been the fact that many countries have done well in a few extremely narrow product segments, such as India's software industry and Bangladesh's success in exporting shirts, trousers and hats to the United States, while Pakistan does well in bed linen and footballs.

So the story is that sectors will relocate, but that this relocation will be 'lumpy', sectorally and in aggregate, with some countries being left out. A corollary of this is that small initial differences – the factors that first attract a sector to a country – will generate large differences in outcomes.



We can think of world economic geography as a balance between concentration forces and dispersion forces

Once again, this points to the importance of policy. Bad policy environments can ensure that a country is left out. Creating a good business environment, institutionally and in terms of infrastructure provision, is essential. The role of pro-active industrial policy remains intensely controversial.

### *Spatial concentration: regions and cities*

Finally, what shifts in economic geography are occurring within countries? High-income countries have an established city structure, but are nevertheless witnessing some changes. After decades of decline, cities are undergoing a renaissance as more knowledge-based activities seek to benefit from clustering.

The situation is more fluid in developing countries, experiencing rapid structural change and migration. Spatial inequality tends to increase during development, often arising from spatial concentration in manufacturing. For example, states in southern India have come to prominence in manufacturing, and Mexican manufacturing has concentrated in regions on the US border, leading to large increases in spatial variation of per capita incomes.

While increasing spatial disparities are a problem for some developing countries, managing the process of urbanisation is a problem for almost all of them. The number of cities in the world with a population of more than one million went from 115 in 1960 to 416 in 2000; for cities of more than four million, the increase was from 18 to 53; and for those with more than 12 million, from one to 11. This indicates that, despite the massive diseconomies associated with developing country mega-cities, there are even more powerful economies of scale, making it worthwhile for firms to locate in these cities.

This creates another major policy challenge. Mega-cities may expand far beyond their efficient scale, but the clustering forces make it difficult for new urban areas to compete and become established. There is a case for policy intervention to decentralise activity, but we remain woefully ignorant about what works and what doesn't.

## Conclusions

There are many reasons for variation in the prosperity of countries and regions. Some factors are truly exogenous – first nature geography – and others are a function of political and institutional history. On top of these exogenous factors, we need to place a theory of the location of economic activity.

International trade theory gets us part of the way, and the new economic geography approach broadens this out to capture (in a micro-founded and evidence-based way) 'endogenous' variations in productivity. The approach offers an explanation of the emergence of disparities between countries and regions – and of their persistence.

It suggests that even as globalisation causes dispersion of activity, so economic development will be in sequence, not in parallel: some countries will experience rapid growth while others will be left behind. At the micro-level, it points to the importance of overcoming coordination failures and threshold effects in growing new cities and in establishing new industries in developing economies.

As  
globalisation  
causes  
dispersion of  
activity, some  
countries will  
experience  
rapid growth  
while others  
will be left  
behind

## Further reading

John Luke Gallup and Jeffrey Sachs (1999), 'Geography and Economic Development', in *Annual World Bank Conference on Development Economics 1998*, World Bank.

Gene Grossman and Esteban Rossi-Hansberg (2006), 'The Rise of Offshoring: It's Not Wine for Cloth Anymore', paper presented at the symposium on 'The New Economic Geography: Effects and Policy Implications', sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming, 24-26 August 2006.

James Harrigan and Anthony J Venables (2006), 'Timeliness and Agglomeration', *Journal of Urban Economics* 59: 300-16.

Paul Krugman and Anthony J Venables (1995), 'Globalisation and the Inequality of Nations', *Quarterly Journal of Economics* 110: 857-80.

Angus Maddison (2001), 'The World Economy: A Millennial Perspective', OECD.

Stephen Redding and Anthony J Venables (2004), 'Economic Geography and International Inequality', *Journal of International Economics* 62: 53-82.

Patricia Rice, Anthony J Venables and Eleonora Patacchini (2006) 'Spatial Determinants of Productivity: Analysis for the UK Regions', *Regional Science and Urban Economics* 36: 727-52.

Stuart Rosenthal and William Strange (2004), 'Evidence on the Nature and Sources of Agglomeration Economies', in *Handbook of Urban and Regional Economics* Vol. 4, North Holland.

This is an edited version of 'Shifts in Economic Geography and their Causes' by Anthony J Venables, CEP Discussion Paper No. 767 (<http://cep.lse.ac.uk/pubs/download/dp0767.pdf>). The paper was originally prepared for the symposium on 'The New Economic Geography: Effects and Policy Implications', sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming, 24-26 August 2006 (<http://www.kc.frb.org/publicat/sympos/2006/sym06prg.htm>).

**Tony Venables** is professor of international economics at LSE, chief economist at the UK government's Department for International Development and former director of CEP's research programme on globalisation.

