

JRS conference paper/EARLY DRAFT  
No citations, not for quotation!

## **Agglomeration, Trade, and Growth: A “History of the World” Point of View**

Michael Storper  
Sciences Po, LSE, UCLA  
[m.storper@lse.ac.uk](mailto:m.storper@lse.ac.uk)  
[storper@ucla.edu](mailto:storper@ucla.edu)  
[michael.storper@sciences-po.fr](mailto:michael.storper@sciences-po.fr)

Paper contributed to the Conference 50<sup>th</sup> anniversary of JRS,  
New York, 23-24 April 2009

### **0. Introduction**

Economic geography, regional science, regional economics, and all forms of social science that deal with space, cities, regions, geography and territory have a checkered history because it is inherently difficult to establish a clear theoretical basis for analyzing a noisy phenomenon such as spatial economic development.<sup>1</sup> This problem seems to have been resolved, for economists at least, in the way the Dixit-Stiglitz-Krugman (DSK) model (Krugman, 1991) overcomes the spatial impossibility theorem in a fully competitive equilibrium. It is responsible for a major expansion of work on geography and trade. The field has now hit its stride through this expansion of effort, with many contributions going well beyond the original DSK-type model.

In spite of many departures from the original model, the field’s legitimacy in economics ultimately comes from reference to the DSK model’s micro-foundations

---

<sup>1</sup> Henceforth I am going to use the term “economic geography” or EG to refer to the whole field, including “regional science.”

and its generality. Economists have rightly reproached geographers and other students of space, no matter how interesting their work may have been, for failing to have a general foundational model. Without some kind of over-arching framework, there is a tendency for partial approaches or empirical investigations to veer off into improbability or irrelevance. Unlike many geographers, I therefore take it as necessary that a field have some kind of general foundations and for that reason I have been enthusiastic about working with NEG-ers.

However, there's a parallel risk in having a general framework: if it contains fundamental errors or gaps or unreasonable assumptions, then much of the partial work that uses it as a reference point will suffer. This is indeed the risk that hovers over the field of EG-NEG-RS today. The field has impressive achievements in the areas of theorizing and measuring trade costs; in incorporating certain kinds of scale economies; and, by incorporating labor mobility, in showing the possibility of shifts, tipping points, and history in patterns of spatial-economic development. Nonetheless, it has largely become devoted to various kinds of comparative statics that integrate these factors. It is, for the most part, ignoring the essential insights of earlier scholars, such as Myrdal, Hirschman and Perroux, that spatial-development is driven by growth, innovation and development : it is fundamentally only explainable in an explicitly dynamic framework . More surprisingly, it largely ignores contemporary economics of endogenous growth. My argument will be that, armed with its progress on trade costs, scale and labor mobility, it now needs to dig deeper and try to create a general framework that will do better than DSK in

orienting its future work, and that such a framework has to be about dynamics and growth.

Economists have made this point recently, among whom I would cite Fujita and Thisse (2002), Baldwin and Martin (2004) and, in another vein, Krugman and Venables (1995), as well as many economic geographers, including myself (Storper and Walker, 1989; Storper, 1997) and Allen Scott (Scott-----), in addition to numerous economic historians and historical geographers. I offer my own twist on this theme, while acknowledging that it is part of this wider flow of ideas.

I will make this argument in four stages. First, I will argue that the many criticisms of the DSK model that are common currency in the RS literature may be valid, but they are largely not the essential ones, because they concern mostly technical details of the Krugman model, and do not address the more fundamental, substantive problems with it. Two of these areas are of particular interest: the foundations of agglomeration economies, and the foundations of cities. My second argument is that these new partial approaches are at risk of losing sight of the most fundamental general, dynamic problems, and indeed display forms of agnosticism about fundamental causal processes and developmental dynamics that risk losing sight of the principal issues. In other words, the field is see-sawing between the reference to a general theory, on one hand (for legitimacy), and putting its nose into details and losing sight of general, fundamental spatial development processes.

Third, the DSK model and a few other attempts at the big picture are basically wrong about sequencing (dynamics of development) and in turn wrong about how equilibrium is ultimately reached. These limitations of the DSK-inspired

general theory are showing up in disturbingly conventional policy applications of the field, that are not unlike old-fashioned regional science.

The fourth proposal I shall make is that the field therefore needs to think about its general orientations once again. From a geographer's perspective, there are elements of this alternative History of the World<sup>2</sup> that could be assembled, and that these could serve as a way to tie together a research agenda that is more about realistic dynamics of growth and spatial-economic development and change.

### **1. The founding model and the origins of spatial concentration**

A great deal has been said about the Dixit-Stiglitz-Krugman model and its basic extensions (**cites**), of which only a few major points need be repeated for present purposes.

The model's principal achievement, as is well known, is to be able to model spatial concentration (and lack thereof) as an endogenous outcome of the economic process, by incorporating labor mobility and assumptions about trade costs, within a variety-based monopolistic competition framework, while leaving general equilibrium intact through assumptions about the size of the agricultural (non-mobile) sector of the economy. The reason it is a real achievement is that it gives researchers the courage to believe that spatial concentration is deeply rooted in the workings of a modern economy, and it incorporates three important realistic phenomena, i.e. labor mobility, product-variety, and firm-level economies of scale.

---

<sup>2</sup> We could also call it, somewhat tongue-in-cheek, a Big Unified Theory, or something like that.

There were forerunner core-periphery formulations in the field of geography and development studies – Myrdal, Hirschman, Perroux, Dixon and Thirlwall – but they never put it all together as in the DSK formulation. The claim is correct that earlier approaches had to make assumptions about specifications in order to get their results (e.g. in Myrdal about total factor productivity, in Hirschman and Perroux about “linkages” instead of the more general-sounding iceberg trade costs used by Krugman; or about the pattern of imports and exports resulting from all this in Dixon-Thirlwall extensions). In some ways, these are differences of order, rather than fundamental nature, since the DSK model is also not solvable without playing with specifications. But certainly, getting beneath linkages and trade to trade costs, is a fundamental break-through, as is the ability to place it all in a general equilibrium framework, all of which liberate the subject for standard treatment.

The main criticisms of the DSK model from the economists have been, curiously (at least to a geographer), mainly technical, centering on its analytical insolvability; the use of iceberg trade cost assumptions; the lack of spatial interaction in competition; the way that concentration scale is derived mechanically from elasticity of substitution of varieties.

I would make different criticisms of the DSK model, based on what I see as fundamental flaws in its substance. A number of these can be cited.

It’s been little recognized, in my opinion, that the only fundamental difference between DSK and old-fashioned central place models is labor mobility, so that the size of the home market is not fixed (leading to the hierarchical nested-market areas of central place), but instead endogenous, leading to a more powerful

impulse to spatial concentration and making the size of home markets endogenous. But is it a difference of fundamental nature? No. It's based on spatially hierarchical patterns of consumption of differentiated goods with different scales, just like in any central place model. This is not just an academic quibble: what it means is that spatial concentration is thought of in the Krugman model entirely through the causal optic of consumer utilities on the part of workers. It fundamentally commits thinking to a view of the spatial dynamic as driven largely by what Muth (1971) would have formulated as "jobs going to people." The people search for consumer utility, this leads to spatial concentration as firms crowd in, with their scale economies, to serve them, in snowball dynamic. The founding moment is jobs to people, and then there is agnosticism about the rest of the process, with a never-ending dynamic of jobs and people determined by complex elasticities of trade costs and scale economies, with different stable patterns and unstable states.

In fact, the Krugman model has no conception of spatial concentration the way that most people in the field of agglomeration economies think of it, as that of production systems tied together in relations of proximity that involve various kinds of Marshall-Arrow externalities. These are nowhere present in the DSK model. To the extent that there is something resembling a localization phenomenon, it comes through the way that if enough producers crowd into a place (due to the causal behavior of worker utilities driving their migration) and if the trade costs in intermediates are such, then there can be secondary agglomeration of the intermediate goods producers. But this does not have a role in driving the basic dynamic of the space economy in DSK or its offshoots.

What does the DSK type of model help us understand? There are probably strong home market effects in certain types of retailing today, as well as in infrastructure, logistics and entertainment services, and certain types of business services. By contrast, it is unlikely that strong HME effects are present in much manufacturing. Indeed, large-scale tradeable manufactures, both intermediate and final, have sufficiently low trade costs today, at almost any level of product variety, that they are not going to crowd into places and create DSK-type spatial concentration. This is probably one reason why empirical research on HME effects has not turned up convincing results, and it can't be fixed by tinkering with the trade costs in the model (ie replacing the iceberg costs with something else).

But the biggest problem with this way of thinking about spatial development is that it derives agglomeration primarily from worker utilities and their movements for consumption opportunities. The real history of modern spatial economic development suggests the opposite: that agglomerations of producers emerge in certain places and that people move to jobs: this is the origin of the local snowball dynamic, and the big causal process in reshaping the space economy. The DSK model gets this almost entirely backward to the extent it deals with it at all. The field needs a theory about the origins of spatial concentration that is quite different from any possible version of DSK modeling. Moreover, as we shall see, spatial interaction would be at the very center of such a model, shaping both the local economic dynamic and the larger process of moving from one spatial-economic equilibrium to another over wider spaces.

## 2. Trade costs, agglomeration, and urbanization

The field has made enormous strides in dealing with core processes that shape spatial development, notably in theorizing and measuring trade costs, agglomeration, urbanization and trade patterns.

For an economic geographer, the persistence of agglomeration in a world of improving transport and communications technologies, concomitantly with increases in long-distance trade and “trade neighborhoods” are what drive the search for explanation. Duranton and Puga (2004) felicitously capture the major sources of spatial concentration in their monikers of “sharing, matching and learning.”

We probably know most about sharing, and especially the influence of trade costs on the degree of co-location within the input-output system; but where we lack complete knowledge is in the interface between industrial organization dynamics and the *genesis* of unit trade costs (more on this shortly).

Matching within labor markets inherently raises issues of industrial organization. Unless we introduce the degree of stability/instability to the employment contract – at any level of skill – it is impossible to show why firms would not choose to spread themselves out and avoid spatial competition in the employment contract. Only an IO-driven, turnover based approach, such as that found in Jayet (1983) can give respond to this issue. And this means that matching is about complex endogenous dynamics.

Learning almost entirely escapes the field of NEG. There is increasing evidence of localized knowledge spillovers, mostly indirectly observed, and proxies in patent



data and some detailed case studies. Learning/innovation has now been subject to better “paper trail” measurement, but there are many aspects that get observed in detailed qualitative research that defy large-scale data. But explaining them from a trade cost agenda seems difficult, without specific add-ons that involve the economics of knowledge transmission and sharing, as well as perhaps elements from contracting. Still, we are very far from understanding *how* knowledgeable people use spatial concentration in order to be more innovative. One finds increasing numbers of articles that claim that skilled people cluster together in order to raise their productivity (and wages), on the assumption that such clustering enables them to generate more knowledge; but examined closely, this is just an assumption about what people do and why they do it based on working backward from the aggregate observation that skilled people do indeed cluster together. It explains nothing. Storper and Venables (2004) have made a stab at it, as have others, but my guess is that the economic sociologists such as Powell (----) have done better than we have. As shall become apparent shortly, in an innovation-driven conception of economic geography such as mine, this is a very important gap.

If we turn to what we do know, it is probably the case that sharing (local trade in intermediates) is the *weakest* source of agglomeration in developed economies, with labor market matching and some kind of technological/knowledge spillover/learning probably underpinning a lot of agglomeration in higher-income developed world regions. In developing economies, local input-output sharing is visibly more important than in developed world cities, for the most part. This is probably because of what has long been observed by students of industrialization

and urbanization (e.g Williamson), that with an immature infrastructure system, intermediate trade costs are sufficiently high that there is a high level of co-located sharing.

In developed countries, the apparently low and declining level of sharing has led many observers to claim that agglomeration is really about competitive selection. But such reasoning is weak on causality: selection might indeed be an outcome, but it surely doesn't account for firm behavior, as if firms would freely choose high levels of spatial interaction with their competitors.

It's pretty much a mystery why we find certain types of local sharing. In other cases, such as "top of the food chain" activities – Wall Street, Silicon Valley, Hollywood – local sharing is probably attributable to the extremely high levels of specialization and uncertainty, which raise the trade costs. Since these industries don't have very high weight-to-value ratios, however, one can still ask what trade costs would be sufficiently high to merit co-location. . It may be that the basis of what co-located firms share and why they stick together is different from what we are accustomed to looking for. Like its neighbor, learning, such sharing may be about sharing knowledge and "cues", and it may in turn be supported both by fluid local labor markets with high levels of turnover and proximity to other producers (the people who do the "knowledge sharing"), so one micro-source of agglomeration morphs into another. Untangling them for empirical purposes is very complex. I believe it requires being quite daring in the conceptual realm, in this case daring to try and theorize novel combinations of the three basic mechanisms of agglomeration and try to measure them (what some of us would call "untraded

interdependencies” or “context”). This hasn’t happened very much, and in my view the categories of sharing, matching and learning, instructive as they are, contain a potentially high level of “dark matter” that suggests the need for understanding local context.

Undoubtedly, though, most standard trade costs for intermediates have declined greatly, and as a consequence, production systems are increasingly unbundling themselves into different territories, leaving over slices of local activities that don’t look like a Marshall-Perroux growth pole. This is even increasingly in the case in Silicon Valley and Hollywood. This raises two important questions for the field. The first concerns whether what is left over and geographically concentrated can be considered an agglomeration that involves local Marshall-Arrow externalities. The answer seems to be positive in many cases, at least as revealed in local TFP levels. Even in the face of lots of unbundling and de-localization, matching and learning are probably holding these bundles of activity in proximity, and much of what they are doing is operating at the directional end of the commodity chain (innovation, design, “conceptual activity), such that access to context, in some case to the possibility of even intermittent face-to-face contact, involving difficult to observe contracting and coordination costs, is a promising line of explanation.

But this is far from the whole story today. There are declining trade costs for wide swaths of both intermediate and final outputs, in a great deal of manufacturing and increasingly in services. Research has indeed considered trade costs, outsourcing, and geographical unbundling. But I’m surprised by the lack of

attention to causal explanation in much of this literature, where a great deal of effort is devoted to trying to show the economic properties of the outcomes (e.g. productivity in places), or the geographical properties of the outcomes (how much offshoring, etc).

These are important, but they give us no handle on what pushes the spatial economy's de-agglomeration processes forward. I think that the principal direction of causality is from the development of the technology and market structure of the sector to its scale and scope attributes, to its basic internal trade cost structure. There are old models, notably the product cycle, that propose interesting stylized facts about how changes in the organization and market structure drive the overall quantity of external transactions of a sector and their unit costs. In their original formulations, they lacked micro-foundations, but these could now be filled in. Basically, these attributes will drive the sector's internal trade costs, by endogenously selecting the quantity of intermediate trade and its qualities (and notably its sensitivity to distance), as well as – for the most part – the trade costs associated with serving final markets (adjusted for the basic technological characteristics of the good/service in question).

Along these lines, it is striking that there have been only limited advances in how spatial economics and industrial organization economics relate to one another, given that product cycle type dynamics are known to drive the evolution of both in powerful ways. In any standard model, what will happen is that elasticities of everything-against-everything (labor costs, land costs, transport costs, you-name-it costs) will determine the way a sector organizes itself (its division of labor), and as a

result its total quantity of trade, and unit trade costs, and the role of distance. This is true by axiom, but not, I suspect, the way the world really works. The challenge is to identify the directions of causality, i.e. which factors are the dog wagging the tail. I will not be agnostic about this question: the “dogs” in the process of sectoral development is technology and the scale and stability of the market. The evolution of its product and production technologies drives scale and scope, and hence drives the division of labor, and hence drives – endogenously – the quantity of trade in the sector, and this in turn interacts with the transport system to drive unit trade costs and the role of distance in them. Location is part and parcel of the determination of trade costs, but there are stronger directions of causality from technology to organization to trade costs and their geography than the other way around. Major improvements in the transport system will have major long run effects on the overall cost structure, but most of these will be strongly filtered through the proliferation of variety and the organizational consequences of this, which will tend to overwhelm the decentralizing effect of lower basic transport costs with higher unit transactions and complexity, leading to a surprising amount of persistent geographical concentration in spite of dramatically improved transport. This is the much-observed persistence, even reinforcement of the gravity effect and of agglomeration and specialization.

This gets back to an issue raised earlier in this paper: do people follow jobs or jobs follow people? This is of course the question of the basic sources of urbanization. As noted, the DSK model falls mostly on the jobs-follow-people side of things. The view of causality in studying agglomeration and de-agglomeration

that I am proposing here would privilege the opposite point of view. There is a cycle of economic and spatial development, with innovation (very broadly understood) at its origin. The type we are most familiar with is when a technology or leading firm gets established somewhere, dominates a commercial breakthrough, and launches the establishment of a local input-output system, which in breaks away from the pack of competitors, and gets the snowball of recursive effects between localization and home market effects going.<sup>3</sup> A second type of innovation is largely political or institutional. The economic reforms in China created a shock to the world production cost structure. Lower long- distance trade costs followed. Another example of political innovation is that in 1947, the Taft-Hartley Act created two separate labor relations systems in the USA, dramatically lowering labor costs in the South, and encouraging a wave of relocations from the North. This happened well before the Interstate Highway System was built.

Regional sharing, matching and learning are apt to be especially strong following a major innovation. Intermediate trade costs are high, leading to agglomeration, and job creation is high, due to the market-expansion, superprofit driven dynamic of the activity. People move to jobs, kicking off the urban development cycle (whether in new places, or “on top” of existing cities). This is one of the two major ways<sup>4</sup> that new urban centers get created in developed economies, because new technologies open up “windows of locational opportunity,”

---

<sup>3</sup> EG: Fairchild/Shockley in Silicon Valley; or Donald Douglass in LA in the 1920s.

<sup>4</sup> The other is through selective de-agglomeration to greenfield sites, to which we return below.

where the sharing, matching and learning processes of existing places may not give them decisive productivity advantages over new places.

I think that the concept of “windows of locational opportunity” is more consistent with the historical evidence of establishment of specialized agglomerations than the ways that most NEG models derive such centers by experimenting with different specifications for trade costs. Those are indeed intriguing, and they can lead to prediction of big new centers (black holes and such). But fundamentally, this is not how such centers are caused: they are caused when some kind of major innovation takes place.

It follows that sources of agglomerations that change the shape of the space economy are not principally about consumer product variety, as in the DSK or central place models, although that may indeed be a *secondary* outcome of the concentration of population in a place as people follow jobs.<sup>5</sup> Consumption-based concentration is never enough to sustain urban concentration, as is demonstrated by the cases of many older cities that have lost their production agglomerations and are in long-term population decline.

Moreover, this innovation-and-growth approach to agglomeration enables us to think in a new way about the product variety question. Learning in agglomerations, combined with the feedbacks from lower long-distance trade costs (enabling specialized new local products to attain wider markets), can endogenously allow the economy to generate much more variety. This in turn can

---

<sup>5</sup> This is also why the abundant models in the field that try to derive the degree of agglomeration from the elasticity of substitution of varieties are pretty much irrelevant.

“seed” the process of agglomeration over and over because such continuous innovation recreates re-generates higher intermediate trade costs (new or continuously improved products often involve substantively complex knowledge or fluctuating markets). The “snowball” of this effect is not the same one as in the DSK model’s drawing in of consumers, but of the economy’s way of generating variety and complexity due to the way it combines local and long distance inputs in new ways.

The contributions of Fujita and Thisse (2002) and Baldwin and Martin (2004) come to mind on this subject. They emphasize the geographical concentration of skilled workers, who innovate, and whose innovations are sources of the long-term growth dynamic of the economy, and where this primary innovation-driven concentration is supplemented by the HME. And both remark that Krugman and Venables (1995) cannot develop an effective explanation of persistent income divergence unless some kind of geography of endogenous technical progress is included in the model. There are strong echoes of Hirschman and Myrdal here. This step in the right direction, however, still requires that we consider causality more thoroughly. It’s hard to know why skilled workers would crowd into a place without some kind of structured domain of activity that attracts them there. Upstream of their crowding together (which may indeed enhance innovation and agglomeration), what is the origin? I have suggested the concept of a technological rupture or a political innovation. Thus, Silicon Valley could have emerged in a lot of places. There was no particular concentration of skilled workers there prior to the decision of Shockley to move to Menlo Park (to be near his



mother), and subsequently for the Fairchild people to break through in their chip design. Downstream from this, skilled workers' skills were created initially through *learning-by-doing in situ*, and *subsequently* through in-migration.

Technological maturity, as in product cycle models, reverses the forces that lead to agglomeration, altering the scale-scope characteristics of the sector, its internal division of labor, the quantity and unit costs of trade, and its geographical characteristics, leading to de-agglomeration. It is technology, market structure and IO that principally drive this possibility of de-agglomeration. The "where" of de-agglomeration is sectorally-specific (type of product and its associated specific trade costs and market locations) and may be affected by the transport infrastructure, but it generally takes the form of geographical unbundling at different territorial scales. People may once again move to jobs in the places affected, and there will be resulting processes of local economic development and urbanization. Their economic characteristics will be quite different from the cities/places positioned higher up in the innovation hierarchy (more on this in the next section). Both firms and individuals are likely to seek cheaper land and housing, and the two are inter-related: firms seek cheaper land because they are no longer constrained by agglomeration economies to co-locate in an innovative milieu and because the rise in scale of production makes it more land-intensive, while the lowering of unit trade costs allows a wider choice of locations involving cheaper land; individuals choose cheaper land and housing because those employed in these types of production are paid less than those in innovative phases. Once again, the directions of causality are

two-way, but not in the same proportions. The “dog” of production wags the tail of the menu of choices.

Is there a pattern of convergence from all this and some kind of “curve” of spatial-economic inequalities?<sup>6</sup> A great deal of ink has been spilled on this question, referring to both continental (eg North America, Europe) and world scales in recent years. It is a subject dear to international economics and to the legitimacy of opening world markets to trade (and, in theory, if not in practice, to labor mobility). Results are highly sensitive to the geographical units/scales used, and to whether data are population-weighted or not, with the former being nearly intractable because of the lack of standardized geographical accounting units.

But beyond these issues, what strikes me is once again the astonishing agnosticism about possible directions of causality and what we think we are detecting in the convergence/divergence debate. To put it bluntly, there will never be convergence, and there will always be fairly high levels of divergence, within and between countries, within the developed world, and between developed and emergent economies. Nothing in the history of the last few hundred years of economic development suggests otherwise. There will, arguably, be periods where convergence is stronger than in other periods, and vice-versa. There will also be shifts in the precise geographical scales at which convergence and divergence tendencies play themselves out: recently in the USA, for example, divergence has

---

<sup>6</sup> I’m not going to discuss in any detail the assumption used in DSK models, that the variety part of the economy is outweighed by the homogeneous goods (ag) sector, as a way of maintaining equilibrium. It’s just a modeling sleight of hand that bears no relevance, that I can tell, for understanding the real dynamics of convergence/divergence.

accelerated at inter- and intra-metropolitan scales, whereas it was in retreat in earlier decades at those scales, and inter-state convergence appears (at least until the recent economic crisis) to have maintained a steady state. The reason for all of this is that, unlike in the DSK model, or in any of the convergence-divergence models used in international trade theory, a realistic view of the economic geography of production, with innovation and product-cycle dynamics at work, suggests that the ongoing agglomeration-deagglomeration/spatial divisions of labor dynamic will always drive a cycle of divergence-convergence-divergence, and so on.<sup>7</sup> It's unlikely that any level of labor or capital mobility could overcome the strong unevenness of the geography of production, and especially the sharp separation of territories according to their position in the innovation/product quality/skills (and hence incomes) hierarchy.

One of the strengths of certain branches of NEG has been, in this vein, to re-open the policy-important question of spatial unevenness in relationship to convergence and "equity." Innovation and growth-based frameworks – by abandoning the idea that there will be convergence, and by incorporating realistic causal drivers into our research – can allow some very promising re-framing of research questions that cut around the standard "convergence is good, divergence is bad" debate, and ask how spatial distribution and aggregate growth are related, as well as about income distribution effects of innovation + agglomeration (Fujita and Thisse, 2002 do this).

---

<sup>7</sup> My position here, using somewhat different reasoning, is quite similar to that of Baldwin and Martin (2004).

This is not enough, however: if we take this mission to its logical conclusion, fully integrating endogenous growth into a field that looks for causal-developmental forces, then its broad economic assumptions have to change. This is our next subject.

### **What kind of economics does spatial economics require?**

One of the great strengths of the DSK model is that it made NEG legitimate by finding a general theoretical solution to the existence of spatial concentration and core-periphery formation. It did so by emphasizing intra-firm economies of scale and certain kinds of pecuniary externalities, and reconciling it all to general equilibrium.

However, most research on spatial concentration has a strong dose of what are commonly labeled “MAR externalities,” a term used by many to capture *external* increasing returns thought to underlie localization and urbanization. It bears repeating that there are no Marshallian externalities in the DSK model. I preface the critical perspective that follows by noting that an immense amount of work on measuring agglomeration economies has been done by economists recently that we geographers should admire.

This has introduced a strange cleavage in our field, between the DSK type equilibrium, and the states of the world that result from admitting the existence of strongly positive locational externalities due to spatial interdependence. Let’s join this notion to that adumbrated in the previous section, i.e. that the origin of

agglomerations is innovation. The implication of this is that the core dynamic of the spatial economy is generated by powerful *dis*-equilibrium states whose economic expression is localization externalities. Moreover, though any particular such state may be transitional (as we shall see), their general existence is central to the growth of the economy in the long run.

To understand this, consider the problem with the term “MAR externalities.”<sup>8</sup> In the Romer endogenous growth model, increasing returns and general equilibrium are reconciled through the way that technological progress is ultimately non-excludable and non-rival. The possibility of a never-ending re-use of new knowledge creates the growth-inducing increasing return, but non-excludability and non-rivalry spread this use out and subject it to competition dynamics that *erase technological monopoly rents*.

It can readily be seen, then, that Marshall-Arrow localization externalities could not be in the same class of phenomena as Romer externalities. M-A externalities (sharing and learning-related) give rise to *locally-bounded external* increasing returns, because the participants in the “secrets” that are in the local “air” have a restricted entry ticket to these secrets. This access allows there to be *localized knowledge rents*, which in turn are what fund the rapid development of innovation and the expansion dynamic in innovative parts of the economy. This is the principal reason why incomes can have such high levels in certain agglomerations in the world, and get reflected in local wages, land prices, and so on, but also why this can continue for as long as the monopoly M-A rents are being

---

<sup>8</sup> This is a short summary of points I make in greater detail in my Roepke Lecture publication in *Economic Geography*, January 2009.

earned in that area. It explains why people (especially the skilled) follow jobs to these places. So, to put it polemically, the term “MAR externalities” is technically incorrect and should be replaced by the term “MA externalities.”

Of course, if the story ended there, the possibility for economic growth would be very limited, as most of the benefits of new knowledge were bid away by the rent-extracting holders of that knowledge. But, as is implied in a very simplified way by the product cycle, this is not what happens. Knowledge that is invented in a restricted circle and enjoyed for a while by a more extended, but still limited circle, ultimately changes form. Novelty becomes codified and standardized and then its rents are bid away. This is where the Romer effect kicks in, because exclusivity is a temporary thing.

Underlying this process of innovation-based motors of long-term growth is the breaking of spatial monopoly effects. As the knowledge is codified, the geographically-restricted communities described by Marshall give way to wider geographical diffusion of the knowledge, more competitive market structures and, of course, deployment of the knowledge through de-agglomeration of its users (or, more aptly, both de-agglomeration and a variety of other processes such as start-ups and imitation and other forms of innovation diffusion). Moreover, this breakdown of spatial monopoly corresponds to the *economic* diffusion of these innovations into other parts of the economy, in ways not envisaged at the outset, thus generating the powerful increasing returns in the economy as a whole, but under competitive conditions.

This means that the appropriate meta-framework for geographical economics has two, intimately-related, moments: the dis-equilibrium states in which the possibility of new wealth levels are generated through innovation; and the spatial-economic diffusion process in which their rents are bid away toward the establishment of a new equilibrium. The interesting part here is not just being able to describe the arrival hall of equilibrium, but the intricate spatio-temporal process from start toward finish, which is what changes the geography of development in interesting ways. The tendency toward a new equilibrium depends directly on the way that the system generates its creative destruction, through rent-fuelled innovation. Spatial economics cannot do its job if it is constrained to the obsessive modeling of equilibrium states, even if the outlines of those states do constitute a principal tendency of the second half of the spatial-economic development process. The first half deserves equal status in economic theory and empirics, and its existence (for which we should all be thankful, since it's the principal source of new wealth) is why the spatial-economy is not only a convergence process, but also has divergence at its heart. Once again, it's the spatio-temporal dance of the two that generates a given level divergence (or trend one way or another) at any given moment.

Many interesting aspects of this dance have figured in recent research: how trade cost "shocks" can break down local rents; how snowballs can "melt down," etc. However, the key process in creating snowballs and melting them down is the nexus between innovation's sources and innovation's effects on IO and trade costs and "relationships" and their geography. To the extent that innovation processes keep

creating new forms knowledge that isn't entirely codified, and new products whose market structures are characterized by instability, uncertainty and contractual complexity, then there will be various combinations of high trade costs, intense labor market matching dynamics, and most of all, local "buzz" in the capture and creation of new knowledge about products and processes and in access to key system-organizing agents (e.g. venture capitalists and IP lawyers today in high tech): these are MA externalities that generate rents. These can go into reverse, as we noted in the previous section, bringing about the R externalities for the economy as a whole through spatial and economic innovation diffusion. What surprises me about the field is that it has not focused as much energy on this key relationship of space to economic growth as to modeling various specifications of trade costs and trade cost shocks and such. Again, my sense is that the "tail is wagging the dog" in terms of priorities and promise for results.<sup>9 10</sup>

Lurking behind these questions of general theory are what we search for in looking at the long-term evolution of settlement patterns. In some countries, the

---

<sup>9</sup> There have been some indirect attempts to get at this, in for example, models that deal with the spatial distribution of the "innovation sector." But this reduces the phenomenon to a specific case, whereas innovation/knowledge rents are the general and most important driver of the growth dynamic and major shifts in the landscape of development. As I noted earlier, the phenomenon of "learning" and its geography, as well as local labor market dynamics, still have a large "dark matter" component.

<sup>10</sup> Another debate that could be better informed by general theory has to do with "creation" versus "selection" in agglomerations. Standard reasoning leads to thinking that bigger places have more rigorous selection. But in the view elaborated here, monopolistic/oligopolistic competition in innovation-driven agglomerations has less standard selection pressure, and more "innovation" pressure, whereas when production is in the standardization-deagglomeration part of the cycle, selection pressures are much greater. Once again, each "history of the world" generates very different intuitions about what question to ask.



urban system seems to expand in a regular manner, going up in size but keeping its overall structure. So is there some way to explain these aggregate outcomes? The Zipf model has no underlying substantive theory. I am personally skeptical about its applicability to the many countries with urban primacy, but let's assume for a moment that some kind of aggregate structure is in evidence and grows apparently regularly over time. Cities or regions could be considered organizational units in this aggregate economy, with some playing the role of extensive margin that steers the cost structure back to equilibrium. In this view, Phoenix and Las Vegas are there to tame land and house prices in Boston and San Francisco, and the key causal mechanisms in this process are flows of location-indifferent firms and preference-seeking workers.

But, as should be evident from the preceding discussions, I do not believe that these mechanisms can support the weight that is put on them. If there is a long-term pattern to be explained, it must come from the spatial patterning of agglomerations and how that influences the number and size of cities, as in the inspiration of the original Henderson model on this subject. The question is how the divisibility of productive activities evolves in a way that might give us demographic patterns that have some regularity, as people follow jobs and generate local development dynamics. What exactly is the role of land markets, transport systems, and preference-seeking individuals in generating these patterns of firm de-agglomeration (another way of putting this is: how spatially indifferent are firms at different stages in the development cycle)? What, by contrast, is the role of the endogenously-driven agglomeration mosaic in generating them? What is the result

of policy in rigging these two types of mechanism, especially the former? The good news is that our field has begun to pose these questions in a policy-relevant way, but I would be very suspicious, given what we have said about agglomeration dynamics, of any a priori assumptions about deep long term universal regularities in settlement patterns or their optimal form. In any case, any attempt to theorize overall settlement patterns needs clear theories of causality, and should be explicit about the “history of the world” assumptions they use, and should test alternative possible histories.

Thus, no single equilibrium framework can capture both the innovation-driven growth of the economy and of particular places within it, and the de-agglomeration, innovation diffusion and regularization of many markets. The most promising way to deal with the creation-growth-development side of the economy, and the adjustment-allocation-optimization side of it, is to admit this, as I have suggested in separating MA from R externalities and that they can be re-linked through a framework built around the spatio-temporal process of dynamics and change.

### **What should we concentrate on?**

The foregoing remarks now lead to a vision of a set of broad orientations for the field of EG. There are two main components of this: the first is some key constitutive mechanisms of spatial-economic change; the second are large stylized processes that, taken together, are a sort of History of the World which can be useful in orienting specific empirical investigations.

*Constituent forces.* To begin with, there are a few basic causal forces that shape the evolution of the spatial economy:

1. Innovation, its link to agglomeration and the creation or growth of uneven development, density, and incomes: Innovation is sometimes exogenous (major technological or political change), and in some cases it is endogenous to spatial development (localized learning and such). This is how major change and creative disequilibrium is generated in the spatial economy.
2. The “spatial possibility function:” this unattractive neologism is meant to convey the notion that possible locations, for individuals/households and firms, are framed by previous collective decisions: those that determine the cost of distance, and those that make durable, spatially-interdependent fixed investments (thus, choices are “bundled,” rather than independent as in standard goods markets, and supply is rigged by major investments). In the end, this means that incentives are key, and they are “rigged” by these complex collective processes (this is a positive statement with no intended normative implications).
3. Preferences: I place them after #2 deliberately. Firms and individuals have preferences on which they are prepared to act. For firms these preferences are defined by technology and market structure. For individuals they are defined largely by socialization. For both, the possibility of realizing them depends on past investments and rigged incentives.

4. Adjustments toward equilibrium: with the standardization of technologies, and diffusion of innovations, firms slowly move and they move the economy toward equilibrium market structures. People follow these jobs, and – within the possibility set given to them, attempt to satisfy their preferences, also moving certain markets toward equilibrium.

*History of the World.*

In the first book I wrote in this field (Storper and Walker, 1989), and prior to the basic Krugman NEG model, we proposed four major elements of what we called the process of “geographical industrialization:”

1. *expansion*: as gravity models show, the spatial distribution of world output has been spreading. Is this “caused” by reductions in trade costs? Only in small part, since most trade is neighborhood trade (Leamer and Storper, 2001). Historically, processes of expansion seem to both follow from, and to *precede*, large reductions in trade costs. This means that we need to go deeper into this complex sequence to understand spatial economic dynamics.
2. *Differentiation*: its not just that there are comparative advantages, but – and here this is a point of agreement with Krugman – differentiation is endogenous to the system through the importance of agglomeration. It is unavoidable and it generates a tendency toward divergence that is

renewed through growth and innovation (circular and cumulative causation, once again).

3. *Shifting centers*: innovation creates windows of locational opportunity that can open major new geographical centers of gravity, as in California in the 20<sup>th</sup> century or China in the 21<sup>st</sup>. Trade costs will only partially (if at all) explain the location of these major new centers of gravity of the economy.
4. *Instability*: the aforementioned processes all generate “local” (at different geographical scales, from city to nation) economic instability, and therefore the long-run of a place is framed by its capacity to adjust and change; but adjustment and change in geographical space isn’t all driven by standard price-quantity adjustment models. If this were the case, Detroit -- with its houses now selling for \$500 on ebay – should be gaining population. Or at least Rochester should. Many kinds of adjustment, especially “high road” development, involves getting hold of and renewing agglomeration economies and specialization.

These are some basic big processes that our field should be devoted to explaining.

To conclude, the achievements of the field, especially in the areas of the details of agglomeration economics and trade costs, represent major advances. The danger is that now, these mechanisms will be used in an increasingly ad hoc manner, but with better and better econometrics, while avoiding the big questions about directions of causality and good sense about how the economy works as a developmental dynamic. So we would be back to looking for the keys under

lamppost, rather than the world around it. The recent *World Development Report* (2009) is a good example of this risk. We are instructed that “density, distance, and division” are key *determinants* of wealth, and that getting the right configuration of them requires allowing scale economies, factor mobility and transport systems to do their work. The report has adroitly avoided discussing directions of causality, sequences, and drivers of the spatial-economic hierarchy at any given moment in time. Some of its policy recommendations are blandly correct, in that they are warnings against the standard bad things that are often done (protectionism, mercantilism, rural favoritism). But mostly what strikes this reader is its absence of useful insights about how to foment or manage development. This is the major epistemological challenge to our field. In this paper I have attempted to argue that it can be met by engaging the new insights that have been achieved in the recent generation of RS research with a broader framework whose basic orientation is innovation, growth, and development.

