WHY DOES A CITY GROW?
SPECIALIZATION, HUMAN CAPITAL, OR INSTITUTIONS?

Michael Storper
Institut d’Etudes Politiques de Paris and London School of Economics

July 25, 2008

ABSTRACT

Why does a certain metropolitan area grow more than another? The answer to this question has evaded much of the considerable body of scholarship on the topic. One problem may be that some of the frameworks that drive empirical research in this field tend be based on ad hoc combinations of explanatory factors, ranging from natural climate to business climate to land and labor costs. Theoretical approaches emphasize differences in economic specialization: some activities have higher rates of growth than others, and this translates into divergence in medium-term rates of inter-urban growth and income. But specialization itself needs to be explained. International economics has adopted theoretical frameworks for explaining different growth rates and income levels among countries involving multiple causes and their potentially recursive interactions. Three main forces are at the heart of this literature: specialization, labor force and human capital issues, and institutions. This framework can be fruitfully adapted to the analysis of metropolitan growth and change. The thorniest aspect of doing so is to consider recursive relationships among the three in a dynamic model, where specialization, human capital and institutions are endogenous to the explanation, and where causality can reverse over time in complex sequences. In this paper, we lay out the elements of such an approach and argue that it could serve as the basis for a new generation of research on differences in metropolitan growth processes.

JEL: R11, R58, R23, J24, H73

1 Michael Storper, Centre de Sociologie des Organisations and Master of Public Affairs, Institut d’Etudes Politiques de Paris, 13, rue de l’Université, 75007 Paris France/ michael.storper@sciences-po.fr and Department of Geography, London School of Economics, Houghton Street, London WC2A2AE, UK/ m.storper@lse.ac.uk
Introduction

Theories of metropolitan growth and development, despite their long and venerable tradition, have lagged behind the major recent advances in theories of economic growth in general. One reason for this is that urban economies\(^2\) are extremely open, so that they are strongly influenced by national growth and national institutions. Even with this, however, growth and development processes for cities are highly uneven; in most countries, growth levels and per capita incomes exhibit strong and persistent differences across metropolitan areas. Moreover, even in countries with the highest levels of internal factor mobility, there are strong differences in the composition of activity among cities (Glaeser, Scheinkman and Schleifer, 1995; Drennan et al 1996), which seem to generate income differences. This probably means that convergence models of city growth face the same limited match to the facts as do convergence models of international economic growth.

Whereas in international economics, the limitations to convergence have stimulated a fruitful theoretical debate about the forces for non-convergence and differentiation, and in urban economics, there is a smaller literature devoted to this problem (Sala-i-Martin, 2002; Barro, 1996; Trefler, 1993).\(^3\) The bulk of studies of why some city-regions have higher income levels, or more overall growth than others, employ a method best described as inductive empiricism. The modal study of “why this or that city grew or did not grow” is to measure a list of factors supposed to contribute to growth. At the top of the list is usually economic specialization, measured usually through shift-share analysis. A city grows by getting more of a sector, or when the sector grows faster in the city than at the national level.

\(^2\) In this paper, « urban economy » will be used interchangeably with « metropolitan economy, », « city-region economy » or “cities” to refer to economies of metropolitan areas. An initial version of this paper was presented to the Fudan University International Urban Forum (FUIUF) Shanghai, November 2006.

\(^3\) The recent literature on convergence and divergence of urban incomes includes, notably, Drennan, Tobier, and Lewis, 1996.
Statistical extensions of the specialization theme include export-base models and multipliers. A mix of factors such as labor costs, land costs, regulation, business climates and so on, are then adduced as explanations for why the city does well or poorly via the evolution of its economic base (cf. Glaeser and Shapiro, 2001; Glaeser et al. 1992; Glaeser et al 1995).

The big problem with these studies is that these latter factors cannot be assumed to be causal explanations. There is a correlation between labor and especially land costs and where an activity locates within a metropolitan regions (expensive in the center, cheaper in the “periphery”). There is also a correlation between labor and land costs and which types of activities go to which types of cities, which mirrors international specialization patterns between high-wage/capital intensive places and low-wage labor-intensive places. As such, these costs matter in explaining why New York, San Francisco and London have little durable goods manufacturing, while certain cities in the Deep South of the USA or southern Europe have a lot of it. But there isn’t any relationship to why an activity locates in one particular city or another within a class of structurally-similar cities; for example, they cannot account for why San Francisco has so much more high tech than Los Angeles with this kind of reasoning. And while in general, there is higher productivity in high-cost places, there are no strong inter-urban locational adjustments of shares according to real productivity differences for a given activity – productivity within sectors seems to be endogenous to the places where they are already concentrated (Sveikauskas, 1975). Only for sectors that are deconcentrating does inter-urban productivity comparison seem to make a difference for where they are subsequently located, in a different class of cities. And once again, it cannot tell us which cities in that class will get the activity.

Neoclassical general-equilibrium models of urban economic growth have more theoretical sophistication than the empiricist-accounting type studies, but their assumptions are largely unsuited to explaining urban growth, in general or in particular. The closer they
get to pure general equilibrium modeling, the more they lose the ability to explain urban concentration and specialization – the heart of urban economies (Krugman, 1991a). The fundamental theorem of neoclassical economics is fundamentally incompatible with “second nature” urbanization – i.e. created from within the economy, rather than from accidents of location due to harbors or seacoasts – because there would be no spatial concentration in the seamless world it assumes (Fujita and Thisse, 2002).

Explaining spatial concentration is precisely the strong suit of the one part of urban economics that has made big theoretical progress in recent years, the New Economic Geography (Fujita, 2002; Krugman, 1991a). Moreover, it allows us to show why, say, an extremely expensive high-wage city-region such as San Francisco can continue to grow in high-technology: the factor cost increases are outstripped by the endogeneous increases in efficiency of the activities concentrated there. Specialization is driven by this endogenous formation and acceleration of agglomeration economies (Rosenthal and Strange, 2001).

However, the NEG cannot tell us why, in the first place, any particular metropolitan region gets set along the path of specializing in something – such as high tech in San Francisco or financial services in London or entertainment in Los Angeles – as opposed to another. That’s where urban growth studies tend to fall back on very specific ad hoc explanations (Stanford University for SF, or good weather in LA, for example) or excessively general explanations such as business climate or factor costs. Agglomeration economics can tell us why, once an industry gets launched in a place, it tends to keep growing for a long time, with strong path dependencies. Economists have called this the “history matters” part of the growth process (Krugman, 1991b). But they still cannot tell us about historical origins in one particular place versus another. These are relegated to the domain of accidents or specific detailed sequences that are said to not be amenable to theoretical generalization (Krugman, 1999).
Moreover, even with the considerable new insights from the economics of agglomeration, the medium-to-long-run evolution of urban economies remains mostly beyond our grasp. Even the most successfully specialized urban economies ultimately run into problems, and this is because the industries in which they are specialized ultimately either have no further agglomeration economies, or they de-agglomerate, or their products become technologically obsolete (Norton and Rees, 1979). All urban economies, like their national counterparts -- in the presence of technological change and an open trading regime – are faced with the question of adjustment to change. This adjustment comes essentially through sectoral succession: successfully getting new specializations, or retaining the retainable parts of existing specializations, to compensate the ultimate loss of what they have.

It will not do to try and loop back to fully neoclassical (i.e. general equilibrium spatial economics) explanations of this process, either; such models are closed by claiming that the optimal adjustments will happen, that new or changing activities will go to places according to their relative productivity rankings, and that this will determine how the specializations of cities will evolve through time. Since standard models can’t explain the “why” of agglomeration in the first place, they have little to say about changes in the specializations of specific places over time (Storper and Scott, 1997).

Comparative growth theory, mostly as applied to international growth and development comparisons over the medium- to long-run, has made significant progress in this area. It stresses the long-run adaptive capacity of economies, in relationship to the changes in technology and geography that alter the competition of places for different activities (Rodrik, 2007). Thus, it addresses the question raised above, about how economies sustain or do not sustain growth over some period of time, in the face of structural changes -- the capacity to pull through cycles and renew economic growth, both quantitatively and qualitatively (incomes) (Haussman and Rodrik, 2003; Pritchett, 1997). Failure comes when an urban
region either has a shrinking economy, or when it quantitatively grows but its incomes, in absolute or relative terms, decline. The task then becomes two-fold. First, is there some way to explain the origins of successful specialization among metropolitan regions in a similar structural class of regions? Second, given the inevitability of change in the locational patterns of sectors in such a region’s economic base, why do some places seem to do better at adjustment than others, effectively changing their specializations over time?. What might such an explanatory framework look like?

**Three principal sources of long-run urban development**

As noted, economic geography has potent theories of why and when sectors will geographically concentrate, and why they leave or disperse: these are the theories of agglomeration based on internal trade costs, home market effects and possibly localized technological externalities. In any event, when sectors are concentrated in certain regions, they cause the economies of those regions to be *specialized* in those activities, leaving a strong imprint in terms of the quantity and quality (type of jobs and local expenditures) of growth. As a recent indicator of this, Galbraith and Hale (2004) note that the income gained in just four (out of about three thousand) US counties in the late 1990s is sufficient to account for virtually all of the increase in geographical income inequality in the USA in the 1990s! These counties are, needless to say, the cores of the US high-tech boom. Some analysts of agglomeration economies believe that localization and specialization are forces that, in the medium-term, impede geographical income convergence.

As alluded to above, most standard economic theory is not fully comfortable with the notion that “specialization matters” over the long run at the international level, because it believes in factor proportions adjustments to economies that ultimately “wash out” the importance of sectoral specialization for national incomes (Krugman and Obstfeld, 2001).
But while this may (and we should emphasize the conditional, because it is not firmly established) be true in the very long run, there is also agreement that specialization can differentiate economies in the short run, and that difference (and non-convergence) can be prolonged through a succession of different specializations. Why? Sectors or activities at different points in their developmental cycles are characterized by different factor proportions and changes in them (labor versus capital-intensive, for example) (Trefler, 1993; Norton and Rees, 1979). More importantly, they can have different terms of trade with the rest of the economy, according to whether there is import or export-biased growth in the economies with which the city trades (Krugman and Obstfeld, 2001). Newer or more innovative sectors can earn temporary rents on their outputs, which they can then renew through innovation.

Growing sectors generally have technological frontiers that are pushing outward at a higher-rate than older industries, so the firms in them enjoy more innovation opportunities. Moreover, in sectors with growing overall demand and supply that doesn’t keep up in the short run, not only is there a rent-effect, but it’s difficult for any supplier to have a decisive impact on prices, so the places that specialize don’t undermine their own positions easily, as they do in sectors with easily-expandable supply (growth based on more mature or standardized products). The point is: specialization really does matter because it creates significant rents for places in the medium-run, due to monopolistic competition.  

Cities can also be specialized in a way that makes them poorer than the average, but this is generally because they have specializations that are not based on agglomeration economies (strong endogenous forces of proximity). Localization should be the result of strong endogenous forces of proximity in the economy. Three of these are in the intermediate output structure of sectors: inter-firm transactions; labor pooling; and technological spillovers. A fourth concerns the home market effect of concentrating producers and

---

4 But it does not matter in general equilibrium approaches to economic geography. In those, movements in factor stocks and factor prices, combined with the absence of significant agglomeration economies, will eliminate any possible monopolistic competition effects over space (Anas, Arnott and Small, 1998).
consumers who each maximize the benefits of economies of scale and product variety when trade costs are strongly positive, by concentrating together. All of these could generate rents for their host areas and make them richer than average. But one can imagine also that economic activities that have none of these locational processes find themselves together in a certain city or region simply because it has the right factor supply for that sector (say land, or labor or transportation access). This form of development is not agglomerated specialization (with monopolistic competition), but simply a collections of firms in a single industry. Thus, not all high location quotients indicate true specialization in the sense theorized by the New Economic Geography. Rapidly growing cities may have specificity of their economic bases, but without the advantages of specialization noted above. The fast-growing cities in the US interior West (generally low income, low wage) have few agglomeration economies in high-wage sectors; the slower-growing cities of the Northeast have higher incomes and more specificity based on specialization, with the attendant benefits (Drennan, 2002).^5

Growth theory doesn’t stop with specialization in explaining different economic fates of places. The second major axis of growth theory holds that the reason a given stock of economic resources can produce more and more wealth over time from increases in productivity that come from the application of new and better knowledge to production. In turn, this knowledge is embodied in people, and can be measured as the stock of human capital.\(^6\) This human capital becomes an externality for the economy, because knowledge can be recombined and re-used in many different ways; it tends to have an ever-increasing positive impact on productivity (Romer, 1990). In international comparisons, knowledge can grow because of increases in R&D, because of investments in education, or due to improvements in the incentives to apply and exchange information. But most important are the conditions that allow the re-use and re-combination of knowledge in order to generate a

---

\(^5\) There’s a different debate about whether it is better for a city to be specialized or diversified cities, but it shouldn’t be confused with the problem we are discussing here. Duranton and Puga, 2000.

\(^6\) This is known as the Barro-Lucas-Romer theory of economic growth (Barro, 1996).
non-linear and positive effect in creating new knowledge. There is little agreement about whether such conditions come from the general institutional environment, from the design of the R&D system, or from more general incentive structures.

It has been famously observed by Lucas (1988) that skilled people congregate in expensive cities to be near other skilled people. But the literature is silent as to why some cities do this better than others, giving rise to big differences in human capital and associated income levels. Even though globalization has made international flows of knowledge more and more open, regions in a country such as the US remain much more open than national economies: they have more interregional trade and labor flows than do even the most open national economies. This means that the stock of knowledge of any given city-region is intimately related to national education, R&D, and labor migration between regions. But the region may influence internal choices to stay or leave, as well as who is trained inside the region. Thus, the regional human capital stock at any given point in time is, in part, caused by the regional characteristics that attract, retain, and repel people with different kinds of skills.

But, as we shall see, not much is known about these regional forces.

The third major branch of international growth theory argues that institutions determine long-run economic growth (Rodrik et al, 2004; cf. Glaeser et al, 2004; Acemoglu, Johnson and Robinson 2004). By “institutions” is meant a variety of things, ranging from the ways the formal de jure rules of political institutions affect their efficiency in facilitating economic activity, to what we might call de facto governance, referring to the real, on-the-ground ways that public sector agencies and private sector groups and individuals interact in detailed ways to shape the rules and resources of the economy (North, 2006). Let’s now define more precisely what this theme might mean to the study of metropolitan growth.

Three fundamental areas of institutional performance are at the center of growth theory. The ways that institutions shape the microeconomic environment (including what is
commonly called the “business climate”, covering such things as confidence and the ways it affects transactions, discounting and investment levels; the ways that they shape labor force participation and effort levels (sometimes known as the “social policy environment”); and the ways that they shape problem-solving, which determines how well the economy captures new opportunities or misses them (how it adjusts to changing technologies and competitors). There is no precise institutional blueprint for these features; rather they represent outcomes for which there are many functional equivalents, depending on the context. They are not a formula, but a sense of what institutions for growth actually do to sustain growth (Rodrik, 2007).

International comparisons are easier when it comes to the formal dimensions of institutions, because international borders of sovereign countries are “hard” institutional boundaries. Regions in some countries – mostly centralized ones --share many of the de jure institutions of economic governance, whereas in other, less centralized ones, de jure institutional design from one city-region to another can differ. And they surely differ from one country to another (Djankov et al 2003).

Another aspect of research of comparative international research on institutions is the size of governmental units. Political economists have recently theorized that there are tradeoffs between the efficiencies that can be gained from size of jurisdictions, and the losses generated (Alesina and Spolaore, 2005). The bigger the unit, the more likelihood there is that there are more heterogeneous preferences of the people within it, and hence the likelihood that many of those preferences will get “washed out” in the conflicts and compromises that must take place in big jurisdictions. In international growth studies, the performance of countries can be partially attributed to how successfully they combine the advantages of scale while enjoying sufficient convergence of preferences to be able to make strong decisions that have public support (Alesina and Spolaore, 2005). To my knowledge, there is no existing study of
such differences in the distribution of formal governmental competencies within metropolitan regions, the resulting size structure of jurisdictions, and the consequences of this for the ways metropolitan political processes operate.

Moreover, city-regions appear often to have strong de facto differences in their inherited (de facto) political cultures and forms of political mobilization that contribute to governance outcomes, in the same way that scholars have seen these among countries. Some have longer traditions of intense community action and established patterns of government-business-community cooperation there, while others have much more top-down political cultures, for example (Logan and Molotch, 1987; Molotch, 1976). “Social capital” indices, that measure such patterns of participation, show that there is much more participation in some city-regions than in others, though the meaning of his for political outcomes has not yet been established by research (Putnam, 2000).

In the regional development literature, a great deal has been said about institutions, but usually in a different sense from the growth theory literature. Analysts have been interested in the sector-specific institutions that make a region able to help a particular industry to flourish. This is especially the case with respect to high technology clusters and flexible production networks and “innovation-based” sectors (Saxenian, 1994; Becattini, 1990). Questions about how production networks are coordinated, moral hazards contained, and transactions costs minimized, as well as whether sector-specific public goods are provided, are the object of a vibrant literature (Scott, 1993). When the concerns of these two literatures are brought together, they suggest the interesting question of how the broad institutional structure of a region interacts with its sector-specific institutions, such as business associations, labor market networks, relationships to the educational system of the region, and other formal institutions and actor-networks.
The opposite direction of causality has also been considered: sectoral interests may shape the performance of general political institutions and political markets for ideas and programs in the region. This is the point of much of the urban politics literature (Dahl, 1961; Cox, 1993). Powerful interests shape the choices made about urban development, especially in land use, through the place-based politics of landowner and developer groups (urban growth regimes) (Molotch, 1976). This has been extended to sectoral business elites as well, in the notion that they influence local politics through their ability to influence job creation and hence generate revenue for local and regional governments. In this way, sectional preferences can find broad expression through lobbying, interest-peddling, and other means of dominating the resource-allocation and policy-setting processes.

The conception of institutions that we propose to import from growth theory is broader than in most of the urban politics literature. It asks how both formal rules and the de facto political processes capture, retain, or damage economic development (Persson and Tabellini, 2006; Rodrik et al 2004; Glaeser, et al 2004). Moreover, there should be microeconomic effects of institutions, not merely how they affect the political decision-making and resource allocation processes. These outcomes include the ways institutions mobilize private and public actors, and filter others out (dis-incentives). They may do so through the ways they influence the formation of coalitions and their intentional, strategic problem-solving activities, or their unintentional mobilizing and de-mobilizing effects on dispersed private actors. Though elite processes and deliberate use of public power to extract rents and build things are important, so are complex, dispersed collective action problems such as how actor-networks are formed, supported, and sometimes weakened or eliminated.  

---

7 It is important to note that at the inter-national scale, there is now a body of (imperfect) large-scale empirical testing of different hypotheses about how institutions affect growth and performance (Acemoglu, Johnson and Robinson, 2004; Prezeworski et al, 2000). At the inter-regional scale, however, while there is an enormous literature, it is almost entirely qualitative or case-study based; we lack any systematic evidence that institutions, politics and governance at the regional level actually matter to regional economic performance in the medium- to long-run. This is astonishing, given the political attention, money and effort spent on such issues.
Thus, a first way to pull together these insights from international and comparative growth theory, and attempt a first application to the problem identified -- how to explain why certain metropolitan regions do better at growth than others -- looks like the following:

- **Specialization: composition of the economy** (New Economic Geography)
- **Human capital: immigration, retention, loss** (New Growth Theory)
- **Institutions: formal structures, de facto governance** (Institutionalist Political Economy)
- **Economic growth and performance:** Innovation, wages, per cap income, growth
Interactions: endogenous causes, feedbacks, filters

Taking comparative economics and economic growth theories seriously, as a model for why a city-region grows would require that we do more than merely consider the three forces identified above. Their interactions must also be considered. By interactions we mean what economics calls “endogenous” forces, or what most people would call “chicken and egg” issues. In somewhat more technical terms, this means that each of the three independent variables (causes) identified above can – under some circumstances -- become dependent variables (effects of one another), and that they can do so in more than one cycle, thus reversing the direction of causality more than once over the medium-run.

What causes specialization?

If one city has had more favorable evolution of its specializations than another, is this because its labor force development (human capital) has became progressively more oriented toward certain skills than the other’s, so that these cities attract and sustain different industries? A standard version of this argument is that specialization responds to factor endowments – in this case particular kinds of human capital. In the case of high-skill industries, it is difficult to reconcile abundance to lower relative prices of the abundant factor, however. New Economic Geography models do better on this account, via labor pooling models: search and matching within a large pool of both labor and employers, allows firms to minimize labor hoarding and better manage their total labor costs, while allowing workers to better secure jobs, manage change, and build their career-long skills (Jayet, 1983; Combes and Duranton, 2006).

Alternatively, as is argued in much of the growth theory literature, specialization is the result of institutions, which ‘select’ the environment to favor certain activities over others, in
the long run. Thus, institutions either “capture” favorable opportunities or they fail to do so (“repel” them); and they either allow the economy to adapt to changing external circumstances, or they block adaptation. Institutions can capture more or less favorable specializations (Rodrik, Subramanian, Trebbi, 2004).

Some scholars argue, by contrast, that specialization is largely *accidental*: being in the right place at the right time attracts a sector to a place, and from there, forces described by economic geographers “lock the activity into” the place, through agglomeration economies. In this view, then, the causes of specialization are external or ‘exogenous’ (Scott and Storper, 1997; Krugman, 1991a; Davis and Weinstein, 2002). One can think of a combination of these two latter views: thus, accidents either give or take away initial “seeds” of specialization, but institutions then promote adaptation or fail at it, in which case they drive initially favorable accidents away. This type of continuous adaptation is said to involve the sector-specific institutions mentioned above, that solve the problems of the sector, but may equally require measures that affect a number of sectors or impact the regional environment in general: hence the need for problem-solving coalitions that go beyond what an industry is able to do for itself (Scott, 1993).

Another point to bear in mind is that the potential for specialization is always partially independent of any “intra-regional” forces: it simply has to do with the evolving organizational and trade cost structure of the industry in question. As is suggested in a highly simplified way by product cycle models, there is a moment in the life of some industries when they are reorganized and their internal trade costs decline to the point that no local measures can suffice to maintain their core agglomerations and hence nothing will keep them in expensive places (Norton and Rees, 1979). On the other hand, there are some moments in the organizational life of sectors where regional efficiency can be improved and shares of an
industry retained: however, this is usually through innovation within that industry and still involves shedding a lot of the routine production activity (Saxenian, 1994; Amsden, 1989).

These endogeneity issues with respect to regional specialization can be visualized in the following manner:

What causes skills/human capital stock?

Institutions and specialization might also be causes of the human capital stock of a city-region. Overall, some regions are richer and more educated than others, and they reproduce some of this in situ, but they also reinforce differences through differential migration. In addition to ethnic specificity of foreign immigration, and sometimes overlapping with it, regions attract mixes of domestic and foreign immigrants with different skill levels and qualities (Frey, 1995). Linked to this is the retention of highly-skilled individuals trained in situ. Among regions with excellent research and training institutions, some will retain more of their locally highly trained immigrants than others, and some will attract more of the highly trained from elsewhere than others. Does this ultimately affect their specializations and
innovativeness? Could it be that something about politics and governance in each region systematically selects for different quantities and qualities of immigration?

In the recent urban growth literature, much has been made of the twin notions of “creative cities” and “amenity-based cities.” (Florida, 2002; Glaeser et al 2001). Both are stories of economic development driven by human capital. In the creative cities framework, specialization is driven by the attraction of creative workers (whose main component are highly-educated workers generally, with a high proportion working in high technology and finance). These workers in turn are said to accumulate in places because of the amenity of “tolerance,” which is operationalized through the composite variable “diversity.” The “amenity city” argument generalizes this to both highly-educated populations (high culture amenities and bohemian amenities) and less-educated populations (sun, low density). The problem with human capital-driven regional economic growth models is that skilled people appear in most cases to precede the creation of amenities, not principally to follow them. Moreover, successful cities are so heterogeneous in terms of their amenities (Shanghai versus Atlanta, Boston versus Orlando), that the notion of amenities easily becomes a vacuous tautology as it is stretched over more and more heterogeneous cases. Household preferences for residential amenities do seem to drive specific locational choices within metropolitan regions, but not between them, so they generate intra-metropolitan sorting but not regional development as a whole (Cheshire and Sheppard, 1995; Oates, 1969; Hilber and Mayer, 2004).

We earlier noted that labor pooling versions of agglomeration theory can be interpreted as suggesting that big labor pools facilitate specialization (Combes and Duranton, 2006). But it could also work the other way around. If an industry begins to agglomerate in a region, and it has unstable markets due to rapid growth or technological change or product innovation/differentiation, then it is highly desirable for firms to have access to a large pool of
labor, because this makes it easier for them to turnover their labor, by insuring that when
demand grows they can find the kind of labor they need in a short time (Jayet, 1983). Thus,
the clustering together of many such firms with unstable labor demands might generate the
regional labor pool through in-migration or regional learning-by-doing, and then the two
become mutually reinforcing causes (Scott and Storper, 1987). In terms of explaining the
trajectories of places, we are caught in very complex endogeneity dynamics in which the
putatively independent and dominant role of human capital is far from being established.

A picture of these endogeneity issues with respect to labor supply is as follows:

What shapes regional institutional performance?

The possible influence of regional institutions, politics, and policy on economic
development has received a great deal of attention in the literature, but there is little that
operationalizes and measures these relations.  

---

8 One of the most ambitious attempts to date is by Rodriguez Pose, 1998, but he does not directly measure policy outcomes and actions, but rather background proxies for institutions.
The formal (de jure) structure of institutions and hence their formal authority and processes for exercising it is determined by constitutional structure, history, and law. The de facto institutional environment consists of the real behaviors of different groups – governmental, business, community and electoral constituencies. Among the factors that might underlie changes in de facto institutional performance are human capital and specialization. If immigration is strong, and immigrant populations become politically mobilized, for example, they may change the preferences that are expressed through political decisions in areas that influence regional development (education, training, infrastructure, business rules, fiscal policy). Likewise, if economic specializations change, then the business groups that effectively “leave” the region will have weaker voices, and the business groups associated with new or stronger sectors will, presumably, have more influence on decisions affecting development (Cox, 1993; Molotch, 1976; Dahl, 2005). If business and human capital actor-networks that represent powerful specializations in the region mobilize around particular strategies, they may influence the output of regional government institutions in a variety of ways (this is the classical point made by the urban politics/growth machine literature). However, it’s important to note two things for a view of the sources of institutional performance: many different types of actor-networks are likely to be involved; and the ways they get their ideas heard depends in part on the formal structure of institutions. In the metropolitan context, key aspects of this are the degree of fragmentation and the size of units in which politics takes place, as this will affect the costs of achieving consensus and hence the formation of the “market” for ideas and strategies that affect the economy, notably in the domain of problem-solving.

The endogeneity issues with respect to institutions are shown below in simplified form:

---

9 Literature on earlier machine politics deals with this issue (Erie, 1990), but the more recent literature seems to pay less attention to it.
It should by now be clear that all three of the main pillars of a growth theory – the economic geography of production, human capital, and institutions – are likely to be important causes of regional growth and development, but that they may be both independent and intermediate causes of a regional growth trajectory. Only additional detailed theoretical modeling and empirical measurement will determine how much each contributes to the explanation of regional development and growth, and how they interact.

**Improving the analysis of institutions and regional economic performance**

In international comparative growth studies, institutions have taken an increasingly prominent role. At the same time, “institutions” refers to many things, lending a high degree of complexity to methods and results. Politics, institutions, and “governance” have also been important reference points in the literature on cities and regions. Because this is the most complex and heterogeneous of the three main pillars of growth theory, it behoves us now to devote greater attention to institutions in the metropolitan or regional context.
One classical theme in urban growth studies is the influence of central government spending on regional growth. Treated in a simple empirical-descriptive way, this amounts to an “external” or “exogenous” cause of local economic change. In light of the present discussion, such a stance is questionable. Central government spending in regions is generally motivated by political coalitions of regional interests that function in national institutions. Technology is also cited as an exogenous factor. Glaeser and Shapiro (2001) argue that the interstate highway system and air conditioning are key reasons for the development of sprawling, sunny cities in the USA. But this explanation has a difficult time explaining why textile mills already moved south in the 1930s, before the advent of air-conditioning, or why dense urban centers and cold old cities resurged in the 1990s (Storper and Manville, 2006).

These specific issues aside, as argued above, the core issue about institutions is how they might affect long-run specialization of a regional economy, and adjustments of the economic base in the face of technological change, globalization, fragmentation of production chains, and so on. Again as noted, social science does not have a satisfactory answer for why an industry might have a strong cluster in a particular place and not another. It does have fairly good answers for why there will be a clustered rather than dispersed structure in an industry and why that structure undergoes change, as well as why the clusters will be located in a certain general type of economy (developed, high wage, less developed, etc). As with development economics in general, it also lacks fully satisfactory explanations of why some places sometimes change structural class, by moving up or down the hierarchy of development: just as this happens to countries, so it happens to cities (Rodrik, 2007). There is strong reason to believe that appropriately-constructed institutionalist arguments and evidence could advance our ability to explain these processes.
An example will help. In the late 1950s, it was not clear that an area south of San Francisco now known as “Silicon Valley” would become the world center of information technology. The initial pattern of semiconductor production in the USA was quite dispersed, and if there was any nascent geographical center of gravity for the sector, it was the Northeast coast, from Boston to New Jersey. Because the industry was extremely new, it lacked a clear “production process” and “commodity chain,” and therefore did not have clearly-defined “factor demands” or even linkage patterns. However, there were a number of areas in the USA with a lot of engineers working on what would become the new technology, from northern California to southern California to the east coast, and even in the southwest. This is what we previously labeled a “window of locational opportunity” (Scott and Storper, 1987). It had parallels in the late 1920s with the American aircraft industry, or the film industry in the 1900s. It has parallels with nanotechnology and biotechnology today.

The problem is that there are often many places that can satisfy the technical needs of an industry. This may be the case with respect to factor supplies (in the IT example, skilled engineers and inventors), as well as institutions. Some noted analysts of the Silicon Valley case (eg Saxenian, 1994) argue that it was Silicon Valley’s institutions that caused it to capture the industry and that this generated the specialization it still enjoys today as the county with the highest per capita income in the world’s wealthiest economy. But once again, we fall into the endogeneity hole: Boston and Phoenix were also early centers of semiconductor production, and it was not until much later that Silicon Valley really pulled ahead of them. In this case, if it is institutions that caused Silicon Valley’s agglomeration forces to strengthen, the institutions are endogenously formed in place – they emerged after the start of the agglomeration process, as a part of the growth of the industry itself.10

10 This would also be true of the labor force. The labor force could not have “caused” the IT industry to concentrate in Silicon Valley, because in the 1960s, computer engineering was just a loose set of individuals, not a consolidated academic discipline. It is an endogenous outcome of the industry’s development, both as a category of skills and training, and as a geographical concentration of those people in Silicon Valley.
The famous “Stanford business park” story could be considered an example of “institutional entrepreneurship” (by the dean of the Stanford Business School), and hence an institutionalist explanation of why Silicon Valley, as opposed to Boston or Phoenix, got ahead (Saxenian, 2000; Kenney, 2000). In this version of things, Motorola’s massive early investments in Phoenix did not find a favorable institutional environment there, or Motorola simply made the wrong strategic choices, and hence the Phoenix agglomeration folded. But this wouldn’t seem to be the case for Boston, which had multiple and diverse actors who saw what was happening and tried to capture the IT industries at the same time SV did. Saxenian’s (1994) argument that they did not do so very well is convincing, but it does not answer the question of “institutions as chicken” or “institutions as egg.”

If the sector-specific institutions we are referring to here are in large-part created as part of the specialization process, then the “exogenous” force of accident may be said to be at their origins as well. It’s important to understand what this does and does not mean. Perhaps in the Silicon Valley case, there were many regions that were—more or less – equally well-prepared to become the world center of the IT industries. But only one of these “candidate regions” happened to get ahead just a little earlier than the others. This is because some actor in that place came up with a break-through “killer application” that tipped the agglomeration economies toward that place: by taking market share, suppliers streamed into the Valley to fulfill new needs, network efficiencies grew, and other places found themselves out-distanced, even though they were “about equally good as one another” prior to this tipping point. The extreme version of this story of accidents holds that William Shockley, the inventor of the chip, moved to Silicon Valley because he wanted to be closer to his mother, who lived in Menlo Park. This tipped the locational structure of the industry, durably. This kind of “accidents of history” explanation is now incorporated in the New Economic Geography’s basic core-periphery model (Fuchs and Shapira, 2005). In my opinion, it fits well the case of
the aircraft industry in the 1930s, where many localities in the US or elsewhere were equally well-placed to become major centers of aircraft production, but Los Angeles got ahead because of a single event: Donald Douglass invented the DC-3 in Santa Monica, and it “took the market” and tipped the geography of aircraft production there rather than elsewhere (Scott, 1993).

Do these events deserve the analytical status of “fully exogenous accidents?” On the one hand, it is evident that many pre-conditions must be satisfied even to have the possibility of making the right breakthrough and tipping the geography of the sector, capturing the specialization. In this sense, there is a regularity that can be explained by social science. It is a regularity more akin to “climate” than to “today’s weather,” however, and that is the problem. We do not like path dependencies and branching points in social science, because they create a wedge between broad and deep structural conditions and outcomes (Hodgson, 1993). But that may very well be a powerful influence in why some cities grow and develop one way versus another, in the medium-run of thirty-to-fifty years (Boschma and Kloosterman, 2005; Rigby and Essletzbichler, 1997). It is a difficult pill to swallow for those who formulate urban policy and want to predict the outcomes of their efforts and expenditures: there is a chance element in economic development.

On the other hand, the regularities may lie elsewhere. In order to be a candidate city in the first place, certain conditions must be satisfied. In the case of first-mover advantages of the type we are considering here, these conditions are likely to come from institutions, but not of the sector-specific type that have attracted the most attention in the high-tech and cluster literatures. There are still-undiscovered attributes of institutions that seem to prepare city-regions to attract new activities and sustain those that are getting started. These may – indeed should – give rise to sector-specific institutional practices of the type Saxenian documents for Silicon Valley. We do not know enough about these “institutions that capture first-mover
opportunities.” The regularity should emerge when we study these institutions: on average, the places that have them should capture more of whatever “new economy” exists at a given point in time than other places. These places should have institutions that overcome existing problems – including existing interest-group practices for extracting rents, dominating perceptions, or blocking other groups from getting attention in political markets and labor markets.\textsuperscript{11}

Long-term processes of economic development are not, thankfully, entirely dominated by first-mover advantages. There are not enough such advantages to go around. However, there are more opportunities that resemble them than is commonly realized. The economy affords abundant “second-mover” opportunities. As sectors mature, they develop more complex internal divisions of labor, usually leading to the possibility of geographical fragmentation of the sector, and so the initial agglomerations, no matter how powerful, do not stop secondary clusters from emerging. By this time, institutions of places – if they are good at adapting and problem-solving -- can apply more systematic lessons of the past to the process of imitation and capture -- assuming that their political structures allow them to adopt the correct policies. In this way, economic development becomes less arbitrary and accidental. An even more powerful opportunity for second-mover advantages comes through product differentiation and quality ladders (Grossman and Helpman, 1991). Product differentiation and quality ladders are a basis for inter-place differentiation and competition, serving as a formidable opportunity-creating device.

In order to prosper over the medium-run, city-regions need to do more than capture first- and second-mover opportunities. They also need to solve problems in two major areas. On one hand, they may attempt to retain existing activities by continuing to modernize them, and on the other they must cope with loss of activities. Such loss is inevitable when the

\textsuperscript{11} Another example of institutional problem-solving that shapes and shifts the geography of an industry would be the finance sector in Venice-Amsterdam-London-New York from the 16\textsuperscript{th} century onward.
evolution of organization and trade costs in a sector eliminate the options of retaining the industry at realistic ranges of regional factor prices. In this case, sectoral succession through first and second-mover specialization must replace the losses, or the economy moves downward in the hierarchy of incomes and employment.

International development economics suggests that institutions create the conditions within which regional economic actors can engage in this process (Rodrik, 2007). One needs only to look at the success of the Japanese automobile industry today to see how important second-mover strategies can be to the economic geography of development, and how institutions do not transfer readily from place-to-place (Cusumano, 1985; Ellison et al, 2002). The question is whether regions also do the same, since there is a strong national imprint to these institutions. How might institutions shape the “action systems” that seem to underpin specialization processes of capture, imitation, retention and adjustment/succession? If it is not accidents that cause them, then regional institutions may affect specialization through another of our growth theory variables, human capital, but in a very specific sense. The *level* of human capital is not enough. Specialization through the construction of actor networks is very likely driven by the *qualities* of human capital, i.e. how well it is adapted to the specific needs of a sector of activity (Rosen, 1983).

The formation of human capital in a specific sector in the economy is a complex phenomenon that has been analyzed extensively in economic sociology. To simplify the argument: effective performance depends on skills, whether acquired on the job or in educational institutions. But acquiring and using skills also often depends on relationships, i.e. knowing where and with whom to acquire experience and develop further on the basis of existing skills (Granovetter, 1995). Relationships also link individuals to opportunities, through networks. Networks also allow for circulation of talent, and for exchange of

12 In the sense that when Japan entered, they were a second-mover. Subsequently, there is a case to be made that they have become the first-mover.
information that continuously improves skills. Some networks have a strong interpersonal basis, or a combination of impersonal and interpersonal connections between individuals. In other words, “human capital” is really a “networked actor system” for the purposes of this analysis. So, people in the New York financial services cluster generally have very different networks from those in the Silicon Valley information technology cluster. Thus, specialization is driven both by the level and the type of skill.

Labor market networks are “institutions” in and of themselves. They mix formal and informal elements, civil society, and formal governmental institutions (regulation, education, etc). They should figure prominently in any effective account of a specific city’s economic history because they are the fundamental source of the “untraded interdependencies” that underpin agglomeration economies.

The geographies of these networks are only partially understood, and what we do know of them indicates that – like many aspects of the regional economy – they have strong path dependencies, as well as many causes that are not specifically regional. Thus, if we consider high-technology business networks, there are roots in institutions that pre-existed Silicon Valley, such as the military-industrial complex, and the national university and R&D system, as well as private companies in the predecessors to high-tech, such as the radio and television equipment industries. It’s difficult to imagine the perpetuation of specialization in Silicon Valley, New York, Hollywood, Paris or London or Milan without these sector-specific business and labor market networks, deeply intertwined with educational networks, that attract, convey skills to, and retain the people that are key to entrepreneurship and action in a particular industry.

But path dependencies are not everything: Silicon Valley is the exemplar of spatial and organizational rupture with the past. Specialization has to start somewhere, and it requires that we consider the causes/origins of the geography of these networks, as well as
why some effectively solve their problems and go on, while others disintegrate and, with them, the specialization they once supported.

Regional institutions in a broader, less sector-specific sense, should logically have considerable – though perhaps indirect – influence on the formation of these networks, the type and level of human capital in the region, and hence its economic specialization. What might these be? Much of the urban politics literature, as noted, concentrates on rent-extracting behavior by regional elites, their dominance of the political process generally, and especially on their intentions with respect to land development (Molotch, 1976; Logan and Molotch, 1987). But there is relatively little on non-land related growth coalitions. The film industry in Southern California, the financial services industry in New York, and the high tech sector in the San Francisco Bay Area are only secondarily interested in land development. In many countries, by contrast, regional political institutions are often more explicitly oriented toward organizing business activity “from above,” with a goal to furthering their specific ends through a variety of public policies. A lot of ink gets spilled about programs put in place to strengthen particular industry clusters in regions. But in the end, we have little hard evidence on the effect that these regional sector-specific elite strategies have on the formation and geography of these networks, and how they affect the dynamics of capture, retention and adjustment of specializations. Take the example of Silicon Valley: perhaps the institutional entrepreneurs at Stanford, or perhaps accidents, set in motion the creation of agglomeration processes of individuals, which in turn led to the creation of wide and deep networks of innovators, which in turn subsequently supported strengthening the agglomeration, and so on. The problem is that no systematic tests of this type of sequences exist in the literature, at least that are known to this author.

Moreover, even if starting points for networks/human capital/agglomeration lie in unique events or unique individuals, one asks whether there is a wider logic to why these
individuals did what they did where they did it. Did the Dean of the Stanford Business School, for example, go west because in the East he felt unable to be entrepreneurial? In this case, was there something about the institutional environment that was propitious to the application of his talents in a particular place?

Finally, leadership may get something started, but it can wither on the vine if it is not appropriately nourished. There are many examples in the history of innovation where superior ideas do not get implemented because they do not find a favorable environment (Mokyr, 1990; North, 2005) These environments should have some regularities we can understand: institutions, not unique genius or pure individual strategy.

The structure of institutions and the regional policy process

In the preceding discussion, we have concentrated on how economic opportunities may intersect with strategic institutional outputs of regional politics. This opens up the natural question as to why regions “do” different things in this domain. No complete answer could ever be proposed for such a question, but there is one dimension that emerges from the recent institutions and growth literature – and which intersects with a classical literature in urban politics and governance – that can be suggested as an important topic for research. It concerns the geography of political jurisdictions in regions and they ways they may influence voice and coalition-creation in regions, and hence affect policy agendas.

The urban environment is one that differs from “normal” factor markets in an economy: it involves spatial interdependency (externalities) and “bundling” (to locate in a place you take a package of things, which you cannot separate) (Storper and Manville, 2006). This gives rise to one of the principal specific qualities of the urban realm: the tendency for households and business to use locational decisions to optimize their benefits, often by choosing jurisdictions within the fragmented multi-jurisdictional space of the metropolitan
region. These “Tiebout-Schelling” dynamics of choice, often involving a tendency for households and businesses to seek out others who are like them, give rise to a strong mechanism of spatial-sorting or self-segregation, and hence a complex mosaic of difference within the diversified metropolis (Tiebout, 1957; Heikkla, 1996; Kenyon, 1997; Kenyon and Kincaid, 1991; Schelling, 1978).

Smaller jurisdictions are likely to have populations with more homogeneous preferences and lower costs of debate and compromise, allowing more initiatives to see the light of day; however, they sometimes need cross-jurisdictional bridging for regional policies or investments with strongly positive scale economies (Alesina and Spolaore, 2005; Aghion, Alesina and Trebbi, 2005). On the other hand, regions with a centralized institutions may have problems with overly-heterogeneous preferences, but can be very efficient at region-wide decisions. Their challenge is decentralizing a certain part of the decision-making/political process. This reasoning leads to the hypothesis that the “worst of all worlds” in terms of the de jure structure of decision-making would be to have neither the responsiveness that allows new initiatives to come up from the bottom, nor the centralization that allows large-scale regional compromises to be forged (cf. Rose-Ackerman, 1983; Stiglitz, 1983). In systems that combine too much bigness without centralization, the bigness of the units usually blocks initiatives, while fragmentation blocks cooperation among large rival units, or even worse, when there is one unit big enough to go it alone without all the others, but not big enough to achieve region-wide compromises. This is a condition of standoff or blockage:
We can think about this in relation to the principal domains of local-regional “hard” power – land use and public investments, and “soft power” (i.e. more general measures to affect regional business climate and quality of life). The fragmented de jure structure will, all other things being equal, allow for a competition of ideas and approaches among a greater number of more homogeneous jurisdictions. This could allow a number of them to do everything wrong, but will also allow many do things “right” through the grouping together of people who share such preferences -- elites and others. The prospect of severe segregation is always strongly present in such a system, as well as failing to achieve regional coordination where
it’s necessary. But the competition from successful places creates a higher probability that public collective action could suggest the “high road” to other jurisdictions and drag them along (also because successful places generate more tax revenues), and that this will, in turn, create a general “upward path” in the dynamics of regional labor market signaling and attraction. The balance of outcomes is not determined by the de jure structure, but by the interplay of real sorting dynamics and the de facto politics that occurs within these jurisdictionally-sorted units and between them.

At the other extreme, a highly centralized metropolis, where there is a dominant jurisdiction (city or county in the US context), faces different dynamics. On the one hand, it has the heterogeneous preferences problem in its big central jurisdiction; but on the other, when this jurisdiction wants to do something big, it can. Hence, there will be a bias toward certain kinds of big projects, and a tendency for many little ones to get shoved under the rug – unless there are political innovations within the big unit, that combine centralization and decentralization (one thinks of the role of New York City in the NY metro area). It’s also conceivable that a strong elite bias gets displayed, if the hegemonic jurisdiction has powerfully organized elites. These elites can, of course, be more or less intelligent when it comes to foresight and problem-solving, so performance can be highly variable. Land use and public investment decisions can, in any case, be powerfully directed toward strong effects on the conditions for labor market/business network formation and sustenance, and hence on specialization. Centralized metropolitan institutions can also fall prey to predatory elites who use centralization for short-term interests and drive specialization down the wrong pathway.

A more problematic case lies between these extremes, of a metropolis with a de jure institutional structure which is neither fragmented among equals, nor centered on a hegemon. It may contain some big units, but that are not hegemonic --- thus denying it both the advantages of centralization and those of smallness and homogeneous preferences. The risk
is that it stifles the expression of land use and public investment projects that express either “many flowers blooming” as in the first case (and hence creative competition within the metro area) or powerful “big time projects” as in the second. It may be rudderless: Los Angeles comes to mind. The City of Los Angeles is not as proportionally powerful in its metro area as New York in its, but it’s big enough to stop almost any wider regional initiative. Los Angeles County is huge (9.5 million persons) and heterogeneous, but not sufficiently piloted by its central city (3.8 million) for the latter to be able to impose its will on its neighboring counties. LA is neither an elite-dominated metropolis like New York, nor an internally fragmented, competitive-cooperative metropolis like San Francisco (Abu-Lughod, 1999; Jaher, 1982; Keil, 2000). The result is that there is possibly the worst of both worlds, with little possibility for creative initiatives to get support, but with elites that are also disorganized. The low road tends to prevail in this case: with ineffective public investments, increasingly segregated private land use decisions, and no strong incentives to create broad problem-solving coalitions (Purcell, 2000). As even the highly skilled have little regional power, they increasingly self-segregate. In the rest of the regional space, with their withdrawal and the absence of effective coalition action, the door is open to a “low road” of economic development, because of the failure to mobilize resources to drive specialization favorably.¹⁴

**Conclusion**

I have made an admittedly complex argument here, but the subject seems to require complexity. Thus far, using simpler concepts of what makes certain cities grow one way and others another way have not established much that is convincing about specific urban growth pathways. Yet policymakers spend huge amounts of money and attention promising to do

---

¹³ But LA certainly was an elite-dominated metropolis until the 1980s, when there was a tide of community mobilization and fragmentation of the regional elites (Fogelson, 1993).

¹⁴ Thus, whereas the per capita income of the five-county Los Angeles metropolitan area was 93% that of the San Francisco ten county metropolitan area in 1970, in 2004 it was only 68%.
something about them, and suggesting they know how to affect these pathways in a positive way. In light of this considerable gap between political action and scientific knowledge of the subject, serious reconstruction of the frameworks we use to carry out research on metropolitan growth and development may well be merited.

REFERENCES


Fuchs, G and Shapira, P, 2005, eds, Rethinking Regional Innovation and Change: Path Dependency or Regional Breakthrough. Frankfurt: Springer.


Kenyon, D and Kincaid, J, eds, 1991, *Competition among States and Local Governments*


