Working Paper 2004-01

The Spatial Organization of Cities:

Deliberate Outcome or Unforeseen Consequence?

Alain Bertaud

Institute of Urban and Regional Development
University of California at Berkeley

Alain Bertaud is an urban planner with over 30 years of international professional experience in America, Europe, and Asia. He is an expert on land use issues, specializing in the spatial aspects of urban development. Since his retirement from the World Bank in 1999, he has been working as a consultant for the World Bank and for private clients. At the World Bank, he was a Principal Urban Planner in the Urban Development Division. During his 20 years of service with the Bank, Bertaud participated in the design and appraisal of large urban infrastructure and housing projects in Russia, China, Eastern and Central Europe and many countries of East and South Asia. His most recent work involves advising municipalities in land use and land regulatory issues as they relate to land markets.

The Spatial Organization of Cities: Deliberate Outcome or Unforeseen Consequence?

Alain Bertaud

Executive Summary

The raison d'être of large cities is the increasing return to scale inherent in large labor markets. Cities' economic efficiency requires, therefore, avoiding any spatial fragmentation of labor markets. In simpler terms, it means that all the locations where jobs are offered should, at least potentially, be physically accessible from all households' place of residence within about an hour's travel time. This requirement should be born in mind when evaluating alternative urban shapes. Any type of spatial organization implying that homes and jobs should be matched individually—i.e., that workers need to have good access only to their current job location—contradicts our premise that large competitive labor markets are efficient and that this efficiency alone justifies the complexity and high operating costs of large cities.

Spatial indicators allow comparison of cities' structures and monitoring over time of the evolution of individual cities' spatial organization. Urban spatial structures can be defined and compared by using a number of indicators related to average land consumption, the spatial distribution of population and the pattern of daily trips. Comparing the value of these indicators among cities shows amazingly resilient common features—such as the negatively sloped density and land price gradient—but also variations of several orders of magnitude, such as the land consumption per person between Asian and North American cities.

Some spatial structures are more compatible than others with environmental and social objectives. Is it possible to establish linkages between spatial structure and city performance in various sectors? In this paper, we look more particularly at the link between city shape and (i) transit use and motorization, (ii) air pollution due to transport and (iii) poverty. We found that dense, contiguously urbanized, and dominantly monocentric cities are favorable to transit and may significantly reduce trip length, and as a consequence, the total amount of pollutant emitted by transport. However, in the absence of adequate traffic management in the central parts of cities, the concentration of pollution might be higher in dense dominantly monocentric cities. Dense monocentric cities have typically higher land prices and therefore tend to reduce the housing floor space and land consumption of the poor while they tend to provide better and cheaper access to most jobs.

Can urban planners influence a city spatial structure? Should they? Should urban planners attempt to change a city's spatial structure in order to improve a city's performance in a particular sector such as transport, the environment or access to jobs by the poor? The chances to do so are rather limited and they are long range, but they nevertheless exist. A planner has three tools at his disposal to influence city shape: land use regulations, infrastructure investments and taxation. However, to use these tools effectively, clearly established objectives must be formulated by elected officials. Because there is no optimum city shape per se, a city shape can be "improved" only in relation to priority objectives. Priorities, however, may change with time, while cities' shapes are very resilient.

Inadvertent changes in city shape caused by poorly conceived regulations or infrastructure investments are much more common than voluntary shape changes. Planners should conduct an audit of existing regulations to find out if their combined effects on city shape are consistent with the municipal priorities.

Do cities' shapes tend to converge toward a standard spatial organization? Is there a global trend in the evolution of urban spatial structures? The available empirical evidence suggests that large cities tend to become less monocentric, and as a consequence, the share of transit is eroding in most cities of the world despite heavy investment and subsidies. On the other hand, in cities of Europe and Asia, which have deliberate policies to provide adequate services in high density cores and to invest in urban amenities—urban design, new theaters, museums, pedestrian streets, etc.—land prices in the city center tend to increase. This would indicate that the monocentric model is not dead or even dying and that the centers of large cities can provide attractions which cannot be matched in the suburbs. However, city centers in large cities, however prestigious or attractive, contain only a fraction of the total number of jobs. Telecommuting, which theoretically will do away with the need for face-to-face contact for a large variety of urban activities like jobs, shopping, education, and entertainment, has not yet had a marked effect on the structure of any city. It is all the more important to monitor the evolution of city shapes and the spatial aspects of the land market to detect any changes in locational demand due to the information revolution.

The Spatial Organization of Cities: Deliberate Outcome or Unforeseen Consequence?

Alain Bertaud

Introduction

Urban spatial structures are shaped by market forces interacting with regulations, primary infrastructure investments and taxes. They are usually the unintended result of unforeseen consequences of policies and regulations that were designed without any particular spatial concerns. However, different urban spatial organizations perform differently. For instance, some urban shapes are unfavorable to the development of public transport; others tend to increase the efficiency of public transport while reducing residential floor consumption. Urban spatial structures are very resilient and they evolve only very slowly. For this reason, a city's spatial structure significantly reduces the range of available development options.

Because city development objectives change over time, defining an optimum city shape is impossible. However, it is possible to identify the type of city shape that would be consistent with a specific objective. Alas, mayors usually are obliged to pursue several objectives at once. While choices of appropriate trade-offs between conflicting objectives are political—not technical—decisions best left to elected officials, urban planners should constantly monitor the impact that specific policies may have on city shape. They should be aware of the effects of the most common planning tools—land use regulations, infrastructure investments and taxation—on the spatial organization of a city. They should make sure that the urban shape resulting from their actions will be consistent with the objectives set by elected officials.

Urban shapes are path dependent. The spatial structure of large cities evolves very slowly and can evolve only in a few directions. On a large scale, land that has been developed can never be brought back to nature. Planners should therefore have a good understanding of the potentials and liabilities inherent to the current spatial organization of the city in which they work. This paper demonstrates a number of tools and spatial indicators to understand a city's spatial structure and to help formulate its potential for different development objectives.

The growth of large cities is a self-generated phenomenon, never deliberately planned and often actively discouraged.

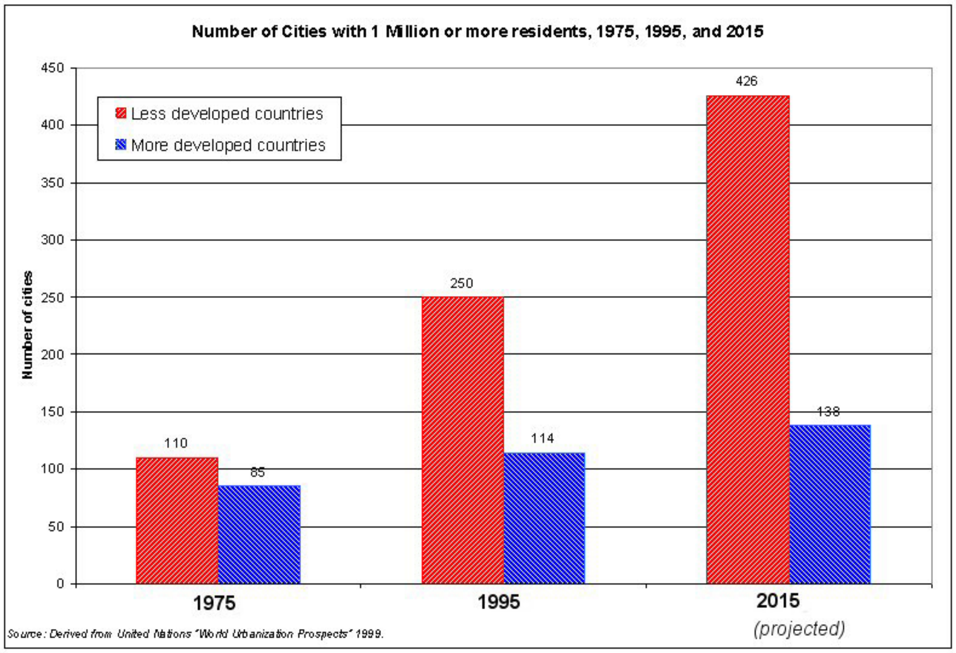
The increase in large metropolitan areas and megacities, occurring worldwide in 2002, has not been stimulated by deliberate policies or even accepted as the unavoidable consequence of economic development.

In the 1960s and 1970s, there was consensus among municipal officials, urban planners, and municipal engineers that the growth of large cities should be limited. The thought was that large cities of several million people would be unmanageable and unlivable. At the same time, there was concern for "geographical equity," an abstract concept implying that all geographical areas of a country should grow at about the same rate. The growth and dominance of large metropolitan areas were considered abnormal, inequitable and malignant in nature.

National urbanization policies tended to promote the growth of small towns and discourage the growth of large urban areas. The consensus on the negative social and economic effects of large cities cut across ideologies, and the same negative bias against large cites could be found in the Soviet Union, Communist China and Cuba, as well as in most market economy countries of Western Europe, Asia and America. The United States seems to have been the only country which escaped the trend, but probably only because the federal government did not have the constitutional right to impose a national policy on city development. On the other hand, states perceived demographic growth as a competitive advantage. It would not have occurred to the state of California to consider its growth inequitable as compared to West Virginia's.

In spite of this universal bias against them, during the last 30 years large cities did grow at a rapid pace, as shown in Figure 1, and are projected to grow further. This is also true for megacities. According to United Nations reports on urbanization, in 1975 less than 2 percent of the global population resided in cities of 10 million or more residents. The proportion now exceeds 4 percent and is projected to top 5 percent by 2015, when almost 400 million people will live in megacities.

Figure 1. Growth of Cities Larger than a Million People Between 1975 and 1995



Source: Derived from United Nations, World Urbanization Prospects, 1999 Revision (2000).

Large labor markets are the only *raison d'être* of large cities.

The fact that large cities have grown and keep growing, in spite of national policies which were biased against them, suggests that some potent economic reasons might be behind this growth. Large cities become more productive than small cities when they can provide larger effective labor markets. A large literature—like Ihlandfeldt (1997) and the classic Goldner (1955)—looks at cities as mainly labor markets, arguing that labor markets have increasing return to scale, which would explain the existence of megacities in spite of the difficulties in managing them. A large unified labor market is the *raison d'être* of large cities. Prud'homme (1996) provides a convincing explanation for the growth of megacities in the last part of the twentieth century: megacities' capacity to maintain a unified labor market is the true long run limit to their size. Market fragmentation due to management or infrastructure failure should therefore result initially in economic decay and eventually in a loss of

population.[1] In this paper, I consider the spatial structure of a city as the possible cause of labor market consolidation or fragmentation. Obviously, the fragmentation of labor markets might have many different other causes—for instance, rigidity of labor laws or racial or sex discrimination.

This paper focuses on the evaluation of the performance of various types of spatial organization. In evaluating spatial structures, we will have to bear in mind that any shape whose effect is to fragment the labor market will not be economically viable in the long run. This is an important reminder, as many planners pretend to solve the logistic problem posed by cities by proposing a spatial organization based on clusters of self-sufficient "urban villages." A viable type of urban structure should therefore allow complete labor mobility within a metropolitan area. Households, whatever their location within the metropolitan area, should be able to reach within a reasonable time (e.g., less than one hour) all the locations where jobs are offered.

Defining urban spatial structures

In order to evaluate the performance of various urban spatial structures, it is necessary to establish some indicators which can be used to measure some of the most important spatial characteristics. Because we aim at an empirical analysis, we will limit ourselves to the indicators which can be easily obtained in most cities by using census data, land use plans and satellite imagery.

To simplify the analysis, we will consider only three aspects of urban spatial structures:

- The pattern of daily trips
- The average built-up density
- Density profile and density gradient

Pattern of daily trips

Traditionally, the monocentric city has been the model most widely used to analyze the spatial organization of cities. The works of Alonso (1964), Muth (1969), and Mills (1972) on density gradients in metropolitan areas are based on the hypothesis of a monocentric city. It has become obvious over the years that the structure of many cities has departed from the monocentric model and that many trip-generating

---

[1] I am certainly not implying here that the quality of infrastructure creates urban growth or that an infrastructure breakdown is the only reason why a city would shrink in size. Exogenous economic factors are, of course, determinant. But if infrastructure is not a sufficient reason to explain growth, the lack of it may explain stagnation despite favorable exogenous economic conditions.

activities are spread in clusters over a wide area outside the traditional Central Business District (CBD). As cities grow in size, the original monocentric structure of large metropolises tends to dissolve progressively into a polycentric structure over time. The CBD loses its primacy, and clusters of activities generating trips spread within the built-up area. Large cities are not born polycentric; they may evolve in that direction. Monocentric and polycentric cities are animals from the same species observed at different times during their evolutionary process. No city is ever 100% monocentric, and it is seldom 100% polycentric (i.e., with no discernable "downtown"). Some cities are dominantly monocentric, others are dominantly polycentric and many are in between. Some circumstances tend to accelerate the mutation toward polycentricity—a historical business center with a low level of amenities, high private-car ownership, cheap land, flat topography, grid street design—and others tend to retard it—a historical center with a high level of amenities, rail-based public transport, radial primary road network, and difficult topography preventing communication between suburbs.

A monocentric city can maintain a unified labor market by providing the capability to move easily along radial roads or rails from the periphery to the center (Figure 2a). The shorter the trip to the CBD, the higher the value of land is. Densities, when market driven, tend to follow the price of land—hence, the negative slope of the density gradient from the center to the periphery.

The growth of polycentric cities is also conditional on providing a unified labor market. Some urban planners idealize polycentric cities by envisioning self-sufficient communities growing around each cluster of employment. According to them, a number of self-sufficient "urban villages" would then aggregate to form a large polycentric metropolis (Figure 2b). In such a large city, trips would be very short; ideally, everybody could even walk or bicycle to work.[2] To my knowledge, no one has ever observed this phenomenon in any large city. A metropolis comprised of self-sufficient "urban villages" would contradict the only valid explanation for the existence and continuous growth of large metropolitan areas: the increasing returns obtained by larger integrated labor markets.[3] The urban village concept is the ultimate labor market fragmentation. Although there are many polycentric cities in the world, there is no known example of an aggregation of small self-sufficient communities. Despite not being encountered in the real world, the utopian

---

2 This is an extreme version of views expressed, for example, by Cervero (1989).

3 Many papers, such as Carlino (1979) and Sveikauskas (1975), document the increasing return to size.

concept of a polycentric city as a cluster of urban villages persists in the minds of many planners. For instance, in some suburbs of Stockholm, urban regulations allow developers to build new dwelling units only to the extent that they can prove there is a corresponding number of jobs in the neighborhood. The satellite towns built around Seoul and Shanghai are another example of the urban village conceit; surveys show that most people living in the new satellite towns commute to work to the main city, while most jobs in the satellite towns are taken by people living in the main city.

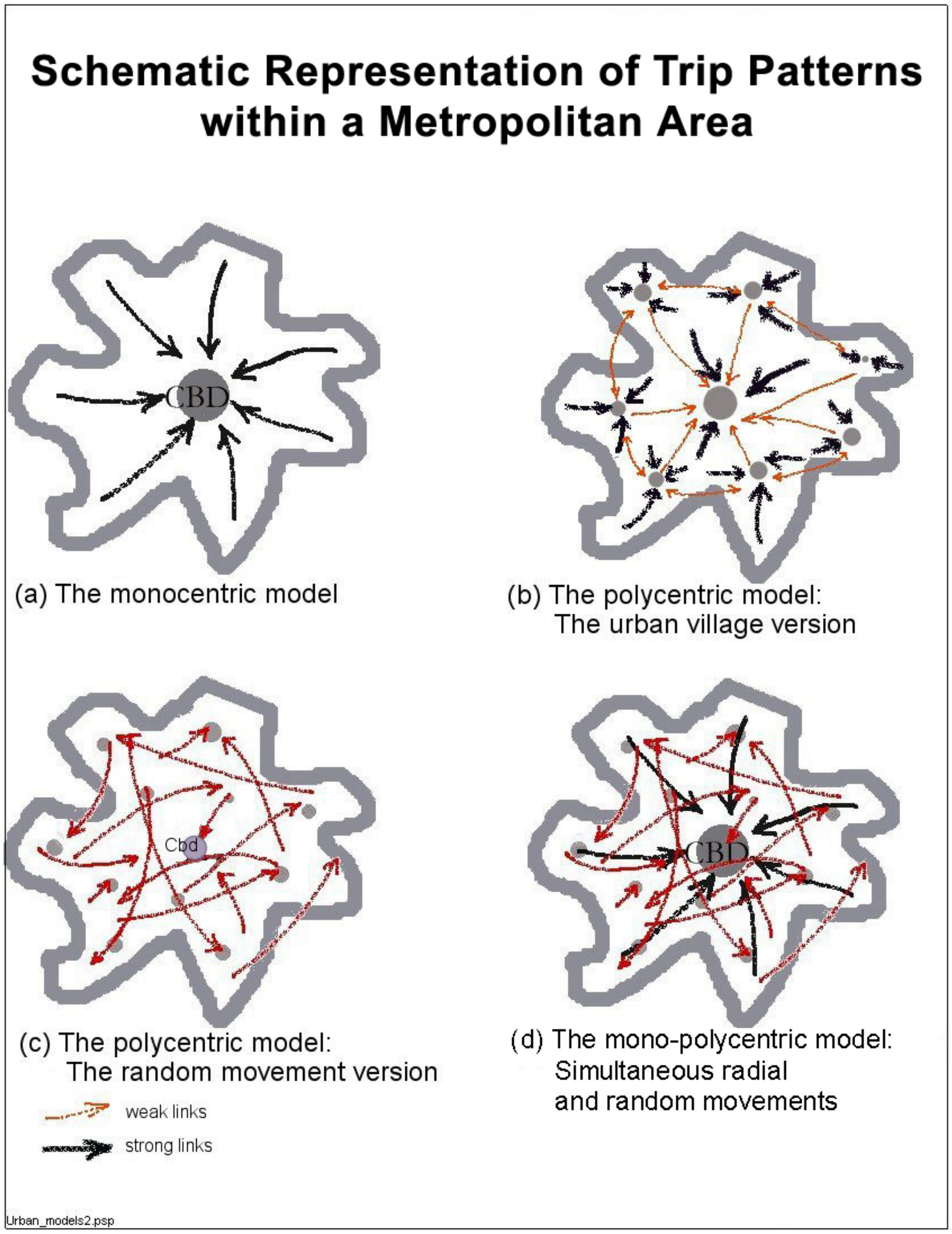
In reality, a polycentric city functions very much in the same way as a monocentric city: jobs, wherever they are, attract people from all over the city. The pattern of trips is different, however. In a polycentric city, each subcenter generates trips from all over the built-up area of the city (Figure 2c). Trips tend to show a wide dispersion of origin and destination, appearing almost random. Trips in a polycentric city will tend to be longer than in a monocentric city, *ceteris paribus*. For a given point in the city, the shorter the sum of trips to all potential destinations, the higher the value of land should be. A geometrically central location will provide trips of a shorter length to all other locations in the city. Therefore, we should expect polycentric cities also to have a negatively sloped density gradient, not necessarily centered on the CBD but on the geometric center of gravity of the urbanized area. The slope of the gradient should be flatter, as the proximity to the center of gravity confers an accessibility advantage that is not as large as in a monocentric city. The existence of a flatter but negatively sloped density gradient in polycentric cities can be observed in cities that are clearly polycentric, like Los Angeles or Atlanta.

Land consumption: Comparative average built-up densities

The amount of land consumed is an important parameter in defining an urban structure. The current concern for "sprawl" is, in fact, a concern for an over-consumption of land by large cities. An accurate, standardized measurement of urban land consumption is indispensable for addressing the issue of sprawl.

Land consumption (area of land per person) is usually measured by its inverse: population density (number of persons per unit of land). Density is often measured as population divided by an administrative boundary—for example, municipal limits. This measure of density is not very useful as municipal limits may include a large amount of vacant land, or even bodies of water. The only way to obtain a meaningful measure of density is to divide population by the built-up area which is consumed by

Figure 2. Pattern of Daily Trips

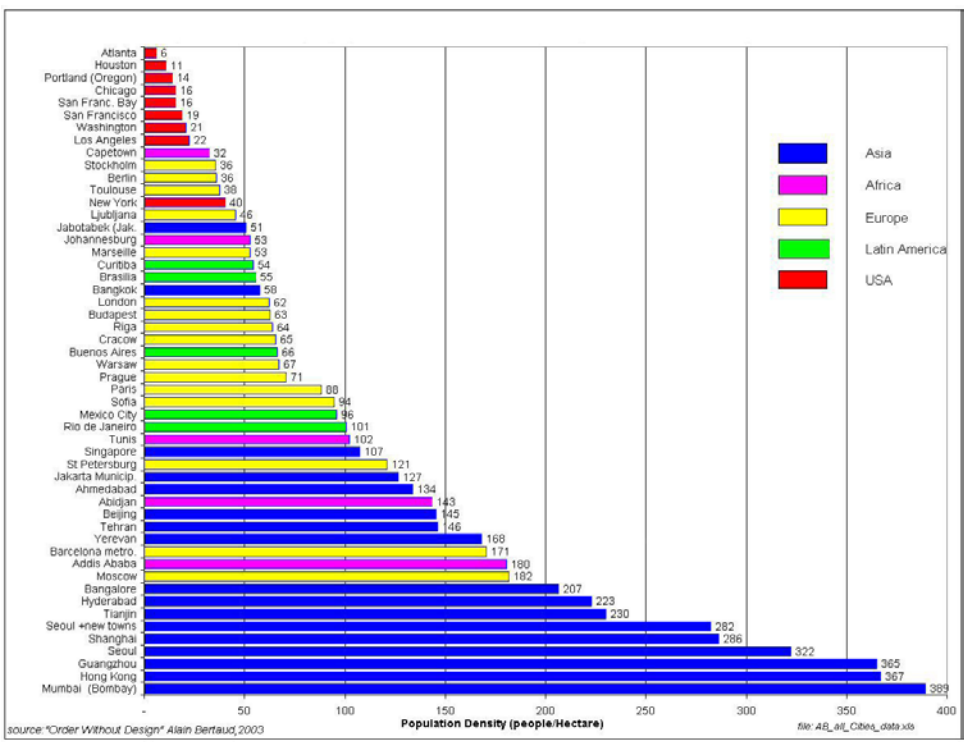


urban activities. The densities mentioned in this paper have all been measured by dividing population by built-up area. Built-up area is defined as including all uses with the exception of contiguous open space larger than 4 hectares, agricultural land, forests, bodies of water and any unused land. In addition, land used by airports and by roads and highways not adjacent to urban used land is not included in the area defined as built-up area.

A comparison between the built-up densities of 49 cities around the world shows differences of several orders of magnitude (Figure 3). One should note that there is no clear correlation between density and income or between density and population size. However, one could make a case for a correlation between the density of a city and its location on a continent. US cities have the lowest densities; African, European, and Latin American cities have medium range densities, while Asian cities have high densities. This may suggest that densities may be strongly influenced by cultural factors. This is not surprising as urban densities are largely influenced by the real estate markets, and therefore, by consumers' trade-offs between commuting distance and land area consumed. The way households make these trade-offs are clearly influenced by culture.

The cities whose densities are shown in Figure 3 are all reasonably successful cities; some might be better managed than others, but the great majority of them constitute the prime economic engine of the country to which they belong. This would suggest that, given the wide range of

Figure 3. Average Population Densities in Built-up Areas in 49 Metropolitan Areas



densities encountered, there is no “right,” “correct,” “manageable,” or “acceptable” range of density per se. None of the cities in the sample shown in Figure 3, representing together about 250 million people,[4] can be said to have too low or too high a density hindering its development or manageability.

Density profiles

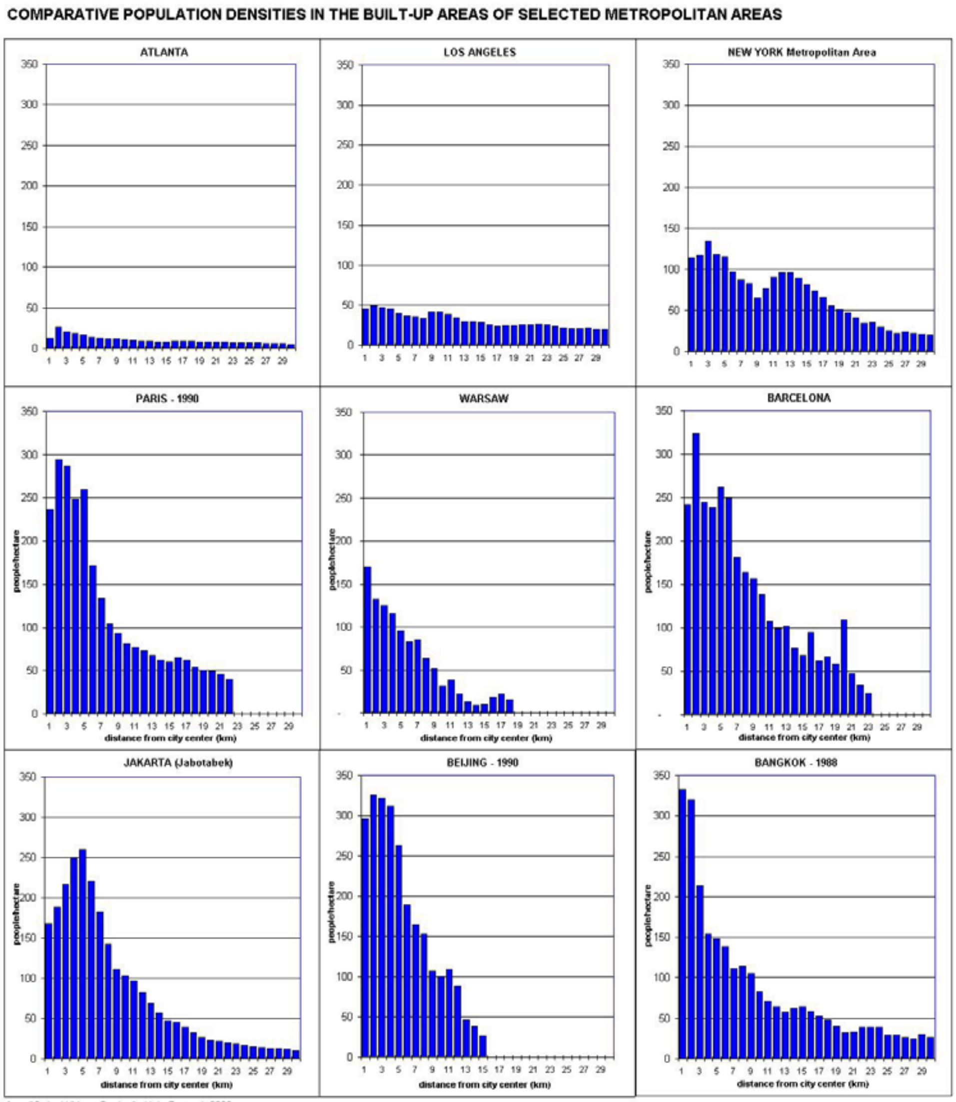
The density profile within a city’s built-up area is a convenient and simplified way to show how the population is distributed within a metropolitan area. Density profiles are based on density maps, which in turn, are based on census data. Census data provide information about the spatial distribution of people when they are at home—for example, between midnight and 6:00AM; census data do not provide any information on where people are during the day. Where people are between midnight and 6:00AM is important because it is the starting point of the daily trips discussed in the section above. What we call the spatial distribution of population is, therefore, an image of the location of the majority of a city’s population between about midnight and six o’clock in the morning. It is important to note that when urban planners show density maps they are, of course, showing the densities around midnight, not the density during the day.

The profile of density provides an image of the distribution of densities by distance from a central point which is usually the center of the central business district (CBD). In the large majority of cities, the density profile follows approximately a negatively sloped exponential curve as predicted by the model developed by Alonso (1964), Mills (1967), and Muth (1969). We can see that this is verified by a sample of 9 cities selected among US, European and Asian cities (Figure 4).

The very large difference in absolute densities around the CBD between US, Asian and European cities can be related to the pattern of daily trips. Dominantly monocentric cities tend to have much higher densities close to the CBD than cities that are dominantly polycentric—such as US cities. The six non-US cities shown in Figure 4 have densities within 4 kilometers of the CBD ranging from 170 to 320 people per hectare (p/ha) compared to a range between 20 p/ha (Atlanta) to 120 p/ha (New York). We will see below that the spatial structures of monocentric high density cities are more compatible with the development of an effective public transport system than those of low density polycentric cities.

[4] Or about 10% of the world’s total urban population in 1990.

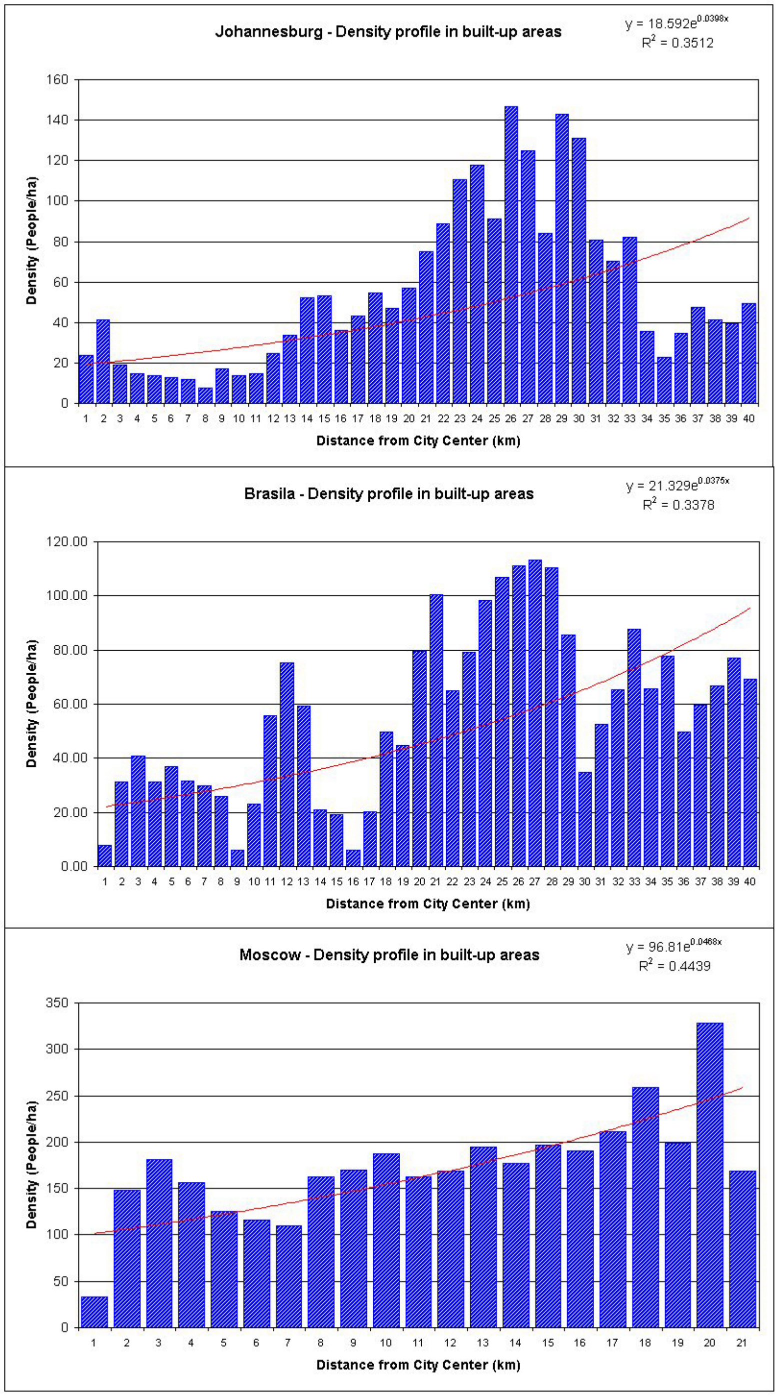
Figure 4. Density Profile of 9 Cities



from "Order Without Design", Alain Bertaud, 2002

The negatively sloped density profile—as seen in Figure 4—is generated by market forces as demonstrated by Alonso, Mills and Muth. This profile is so resilient that even cities with historical interruptions of land markets often show negatively sloped gradients, as do Warsaw and Beijing in Figure 4. However, a few cities where markets were interrupted or absent for long periods show positively sloped gradients (Figure 5).

Figure 5. Brasilia, Johannesburg and Moscow Density Profile



While a high or a low density does not have a necessarily negative effect, a positively sloped density gradient always constitutes a liability for a city. For a given average density, in a city with a positive gradient, the median distance per person to the CBD will always be longer than in an

equivalent city with a negative gradient. It is reasonable to infer that in a city with a positively sloped gradient all trips would be longer.

Moscow, Brasilia and Johannesburg are cities that seem to have very little in common except a history of disturbed land markets. Whether the interruption was caused by Marxist ideology in Moscow, a morbid cult of design in Brasilia or Apartheid in Johannesburg is irrelevant; the spatial outcome is similar. The positively sloped density profiles reveal this common part of their history.

Linkages between spatial structure and transport efficiency

The type of urban structure often defines the most efficient mode of transport. The type of spatial structure—i.e., the degree of monocentricity and density—has a direct impact on trip length, the feasibility of transit being the dominant mode of transport, and pollution.

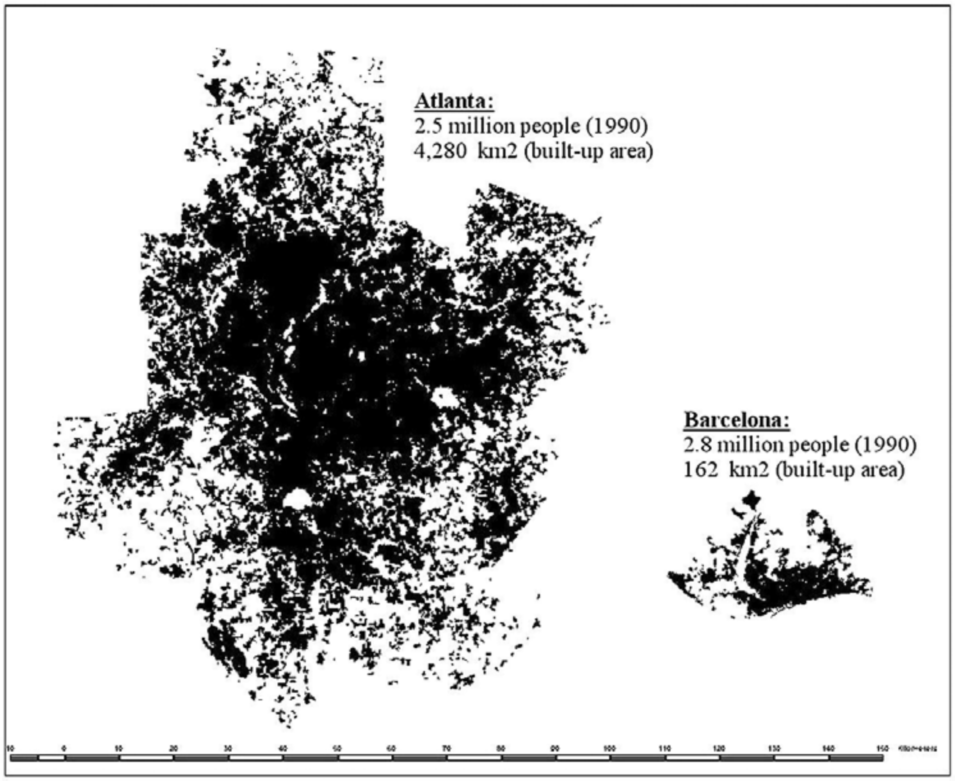
Densities, monocentrism and trip length

For a given population, the higher the density, the smaller the built-up area is. Providing the built-up area is roughly contiguous—i.e., not formed of large isolated areas like satellite towns—trips will be shorter in length in cities with high densities than in cities with low densities. A comparison of the built-up areas of two cities such as Atlanta and Barcelona, with similar populations (about 2.5 million in 1990) but very different average densities, illustrates this point (Figure 6). In Atlanta, the longest possible distance between two points within the built-up area is 137 kilometers; in Barcelona, it is only 37 kilometers. The short trip distance due to high density in Barcelona makes possible a significant number of trips by foot or bicycle. Within the Barcelona municipality, 20% of trips are made by walking. In Atlanta, the number of walking trips is so insignificant that it is not even recorded!

Density, however, is not the only factor that influences trip length. In a dominantly monocentric city, trips usually are shorter as the majority of trips are from the periphery to the CBD. In most dominantly monocentric cities, the population's center of gravity coincides with the CBD, as is the case in New York, London, Paris, Moscow, and Shanghai. In this case, the larger the proportion of trips to the CBD, the shorter the trips will be since, by definition, the center of gravity is the point from which the sum of distance weighted by population is the shortest.

The effect of the spatial distribution of density on trip length is often underestimated. The theoretical graphs in Figure 7 show the large variations in trip length produced by different spatial arrangements for an

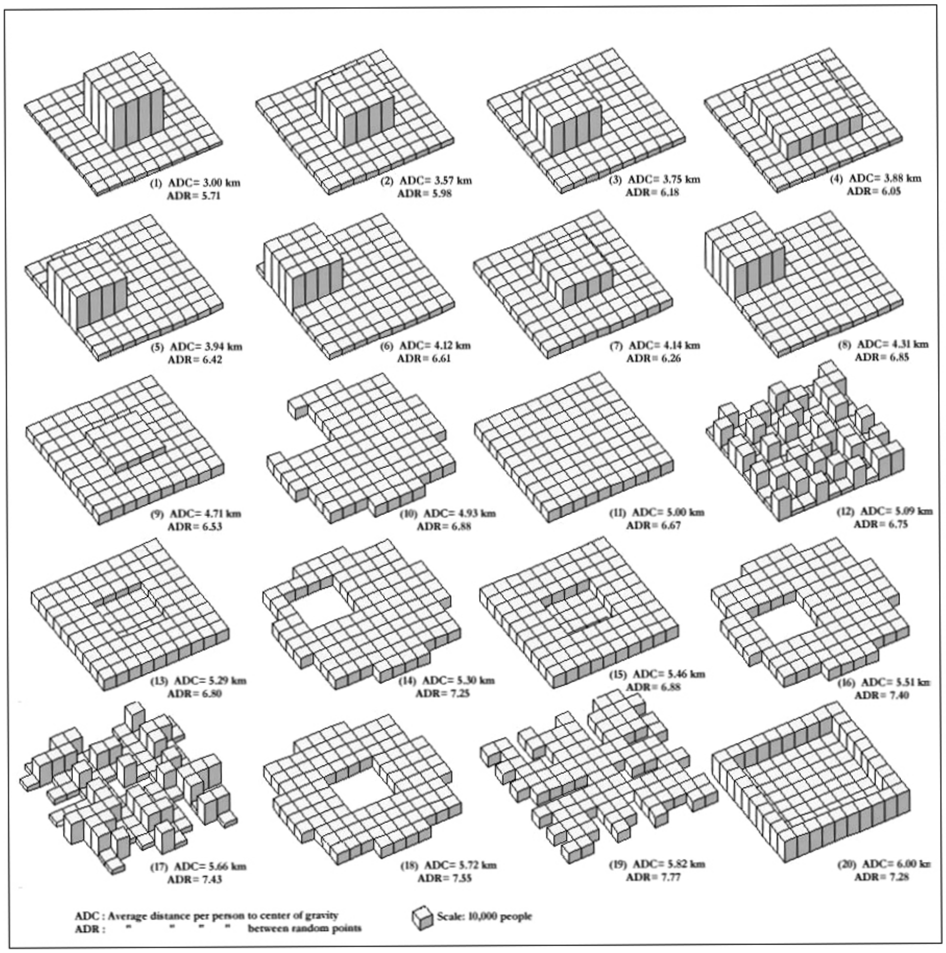
Figure 6. The Built-up Area of Atlanta and Barcelona Represented at the Same Scale



imaginary city whose population and built-up area is kept constant. These variations in trip length occur whether trips are radial or from random origins and destinations.

Let us assume an imaginary city of 1 million people with an average density of 100 people per hectare—i.e., a built-up area of 100 square kilometers. To limit the number of possible shapes, the variations will be limited to those inscribed within a 12-kilometer square. Let us then test the variation of distance per person to the CBD and the average distance per person between random points for 20 variations of typical spatial structures, keeping the average density, population and built-up area constant. The variables are the density of sub-areas, the location of sub-areas with different densities, and the shape of the built-up area within the limit of a 12-kilometer square. The results are shown in Figure 7.

Figure 7. Schematic Representation of Different Distributions of Density in a City with Constant Average Density and Built-up Area



The spatial organization types shown in Figure 7 are presented in order of decreasing performance for average distance to the CBD. The results allow us to make three observations:

1. The variation in performance among types is large. The distance to the CBD doubles between layout 1 and layout 20—from 3 to 6 kilometers, although the shape itself stays inscribed within a square of 12 by 12 kilometers. Between cities of identical average density, the distribution of local densities is a very important factor in determining the length and, therefore, the costs of trips and transport networks.

2. The variation in distance to the CBD is much larger between different spatial arrangements than between the distance to the CBD and the distance between random points for a given shape. Shape itself is

more important in city performance than whether a city is monocentric or polycentric.

3. While a poor performer for the distance to the CBD will generally be a poor performer for average distance between random locations, the correspondence is not linear. Some types of spatial arrangements, which are favorable to monocentric movements, are not favorable to random movements. For instance, the layout ranked 13 for distance to CBD performs better for random movement than the layout ranked 8.

Compatibility of private cars with high densities

In this paper, densities are expressed in persons per hectare of land (p/ha). But they can also be expressed in square meters of land per person. For instance, Atlanta's average density of 6 p/ha corresponds to a land consumption of 1,666 square meters per person, and Barcelona's density of 171 p/ha corresponds to a consumption of 58 square meters per person. In the CBD of the European and Asian cities whose density profiles are shown in Figure 4, the density is around 250 p/ha corresponding to a land consumption of 40 square meters per person.

A private car that moves around and parks in a city needs at least 40 square meters of land space. We can see that in Atlanta a car will occupy only a small fraction of the land available per person, only about 0.4%. By contrast, in the center of an Asian or European city, a private car would require about the same space as a person. The more cars are introduced in the CBD of dense cities, the more they compete for space with people, not only with pedestrians but also commerce, open space and all sorts of amenities. If private cars moving around and parking in a CBD were charged market rents for the space they occupy, the land allocation problem between cars and other activities would be solved. Unfortunately, this is not the case. Many cities have free or subsidized parking along their street curbs, and sometimes even subsidized off-street municipal parking in downtown areas! Because of political difficulties and the high transaction costs of charging market rents for the land occupied by cars in downtown areas, land often must be allocated in an administrative way between cars and other urban activities—hence, the necessity of creating pedestrian-only streets and restricting car access within historical areas, which was successfully done in the historic centers of Cracow and Riga, for instance. The latest experiment in road pricing in the center of the city of London is another example of the difficulty of allocating space efficiently between cars and people.

There is definitely a density threshold beyond which private car access should be severely restricted or even banned. In lower density

areas, low land rents do not justify parking transaction costs; in dense downtown areas with high land rents, having cars pay market rents for the land they occupy is the only way to obtain efficient land allocation. The failure to charge for parking and street space may, in the long run, destroy the amenities of downtown areas because of the misallocation of land among those who pay market rents (i.e., shops, offices, housing) and those who do not (i.e., cars). Singapore is the only city that has attempted to reflect the true costs of introducing cars in its downtown area. The technology the city uses allows the price of downtown access to be finely tuned, while at the same time, keeping the transaction cost low. Ironically, Singapore, by Asian standards, is not particularly dense (Figure 3).

Transit compatibility with various density levels and trip patterns

While we have seen that high densities are incompatible with the use of private cars, the reverse is true for transit. Transit is incompatible with low densities and urban spatial structures that are dominantly polycentric.

There is a purely geometric explanation to why low densities are incompatible with transit. Transit stations and bus stops have to be accessible by people walking to and from their homes or jobs or whatever activity necessitates the trip. A person's walking speed is limited by physiology to about 4.5 kilometers per hour. Acceptable walking time to transit stations varies with culture and income, but surveys show that most people will not walk more than 10 minutes to a transit station or bus stop (although the acceptable walking time is usually higher for a metro station than for a bus stop). A 10-minute walking distance at a speed of 4.5 kph corresponds to a rounded maximum accessibility radius of about 800 meters to a bus stop or metro station. A radius of 800 meters in a street grid pattern will correspond to a catchment area varying between about 110 and 128 hectares depending on transit stop intervals and transit line distances. As a rule of thumb, I use a maximum catchment area of 120 hectares per transit stop. When the number of people living, working or shopping within the 120 hectares falls below a certain threshold, transit becomes impractical to the user and financially infeasible for the operator. There seems to be a consensus among various researchers and operators that the density threshold for transit is around 30 p/ha (Table 1).

The comparison between Atlanta and Barcelona illustrates in a more concrete manner the problems of transit operation in low densities. Atlanta and Barcelona had nearly the same population in 1990—2.5 million and 2.8 million, respectively.

Table 1. Recommended Densities for Transit Operation

Recommended built-up and residential densities for various levels of transit services

	Built-up Density	Residential Density
Boris Pushkarev and Jeffrey Zupan (1982)	p/ha	p/ha
1 Bus: intermed serv, ½ mi. between routes, 40 buses/day 7 du/res ac	29	42
2 Bus: freq serv, ½ mi. between routes, 120 buses/day 15 du/res ac	62	89
3 Light rail: 5 min. peak headways, 9 du/res ac, 25–100 sq. mi. corridor	37	53
4 Rapid tr: 5 min. peak headways, 12 du/res ac, 100–150 sq. mi. corridor	50	71
The Institute of Transportation Engineers (1989) recommends the following minimums:		
1 1 bus/hour, 4–6 du/res ac, 5–8 msf of commercial/office	21	30
2 1 bus/30 min., 7–8 du/res ac, 8–20 msf of commercial/office	31	44
3 Lt rail and feeder buses, 9 du/res ac, 35–50 msf of commercial/office	37	53
Peter Newman and Jeffrey Kenworthy (1989)		
"public transit oriented urban lifestyle"	35	50

Extracted from: **John Holtzclaw, "Using Residential Patterns and Transit to Decrease Auto Dependence and Costs," Natural Resources Defense Council, June 1994.**

1 acre =	0.405 ha
Persons/dw =	2.4
% residential	70%

Barcelona's metro network is 99 kilometers long and 60% of the population lives less than 600 meters from a metro station. Atlanta's metro network is 74 kilometers long—not much different from Barcelona's—but only 4% of the population lives within 800 meters of a metro station. Predictably, in Atlanta only 4.5% of trips are made by transit versus 30% in metropolitan Barcelona.

Suppose that the city of Atlanta wanted to provide its population with the same metro accessibility that Barcelona does (i.e., 60% of the population within 600 meters of a metro station). It would have to build an additional 3,400 kilometers of metro tracks and about 2,800 new metro stations. This enormous new investment would allow the Atlanta metro to potentially transport the same number of people that Barcelona does with only 99 kilometers of tracks and 136 stations.

The example above illustrates the constraint that low density imposes on the operation of transit. Metro track length and stations have been compared, but a comparison between bus line lengths and number of bus stops in Barcelona and Atlanta would give the same results. With its low density of 6 people per hectare—compared to Barcelona's 171 p/ha—

Atlanta would have difficulties developing a viable form of transit—i.e., a transit system that is convenient for the consumer and financially viable for the operator.

In Atlanta's case, the very low density precludes developing transit as an alternative transport to the automobile. "Encouraging" higher density, as many reports are fond of recommending, is not feasible either. To reach the 30 p/ha threshold over a period of 20 years, assuming the historical population growth rate of 2.7% per year continues uninterrupted, the current built-up area would have to shrink by 67%. In other words, about 67% of the existing real estate stock would have to be destroyed; the land over which it lays would have to revert to nature; and the city's population and jobs would have to be moved into the remaining 33% of the city.

The Atlanta example shows how an existing spatial structure constrains the number of alternative strategies available to guide city development. The lack of spatial analysis often leads to the recommendation of infeasible strategies—i.e., strategies which are incompatible with current urban structures.

Density is not the only spatial factor which constrains the development of transit; a dominantly polycentric structure is also a hindrance to transit operation. In monocentric cities, most trips have multiple origins (the suburbs) but have one group of "clustered" destinations (the city center). In polycentric cities, most trips have multiple origins and multiple destinations. Consequently, in a dominantly polycentric city, there is a multiplicity of routes with few riders. As a result, transit systems can operate efficiently in monocentric cities but are difficult to operate in polycentric cities.

Pollution and spatial structure

The amount of air pollution generated by urban transport depends on the length, speed and number of motorized trips and the type of vehicles. For a given urban population, the length and number of motorized daily trips are closely correlated with the average population density in built-up areas and the spatial distribution of trip destinations and origins. Therefore, low density, polycentric types of urbanization have a double effect on pollution generated by transport. First, it increases trip length compared to denser, more monocentric structures, and second, it increases the number of motorized trips as the proportion of transit trips and walking trips decrease with density.

Engine technology and fuel types also play an important role in the amount of vehicular pollution and can counteract or attenuate the effect of

unfavorable spatial structure. The comparison between Atlanta and Barcelona illustrates an interesting example of technology's impact on urban air pollution. In 1999, the average yearly level of nitrogen oxides was 47 $\mu g/m^3$ in Atlanta compared with 55 $\mu g/m^3$ in Barcelona. Air pollution caused by traffic is greater in Barcelona than in Atlanta, despite the fact that Barcelona's density is 28 times higher than Atlanta's, 30% of trips in Barcelona are by transit, and 10% are walking trips. This is due to laxer emission standards for vehicles—in particular, the use of diesel fuel for cars (about 55% of private cars use diesel in Barcelona). In addition, vehicles in Spain tend to be older than those in the US. For the same model, older cars may emit as much as ten times the amount of pollutant emitted by new cars; thus, in some cases, air pollution might be more sensitive to age and quality of vehicles than to urban spatial structure.

Compared to low density polycentric structures, high density monocentric structures certainly tend to decrease the total amount of pollutant emitted by transport. However, the level of pollution exposure in dense monocentric city centers might be higher because of the more intense and slower traffic. Strict bans of on-street parking to increase the speed of traffic flows and general traffic management measures are necessary to decrease high pollution exposure in central city areas.

Spatial structure and poverty

The type of urban spatial structure affects the welfare of the poor in a number of ways. In countries where low-income residents cannot afford individual means of transportation or where the city's expanse precludes walking to work, dense monocentric cities are more favorable to low-income households because they reduce distance and allow an efficient network of public transport.

The poor have greater access to job opportunities in dense, monocentric cities, but land is usually much more expensive there, and as a consequence, low-income households are less able to afford land and floor space for housing than they are in more sprawling polycentric cities.

High density housing requires a much higher quality of infrastructure and urban services than low density housing. For instance, a leaky sewer in a low density settlement (for example, 50 p/ha) does not present as great a health hazard as the same leaky sewer in a dense settlement (for example, 800 p/ha—a common density in residential areas of many Asian cities). In the same way, a deficient solid waste collection system is less damaging to the environment in low density settlements than in high density ones.

To summarize, the poor are better off in a dense monocentric city when it comes to job access; however, in this type of city, they are more likely to consume less land and floor space than in low density polycentric cities, and the quality of their environment might be worse.

Urban laws regulating densities reduce the locational choices of the poor

Low-income households must make trade-offs as do higher income groups when selecting a residential location. Residential mobility—defined as being able to change residential location to maximize one's own welfare—is even more important for low-income households than for higher income groups. However, many well intentioned land use laws and subsidized housing programs tend to severely limit the residential mobility of the poor.

Providing it can be retailed in 'non-lumpy' quantities, urban land is usually affordable for most income groups. For a given land price, each income group adjusts its consumption of land (and therefore density) and makes its own trade-offs between distance to work and land and floor space consumption. However, land use regulations—by establishing minimum plot sizes and floor area ratios—tend to make land consumption lumpy, reducing locational choices of households who can afford only land parcel sizes below the arbitrary legal minimum. Many land use regulations have the effect of segregating the poor in areas which might not be best for their welfare. The theoretical example below illustrates this point.

A city where the land price profile follows the classical model established by Alonso, Muth and Mills is shown in Figure 8 where distance from the center in kilometers is represented horizontally and land price in US dollars is represented on the vertical axis.

Let us assume that two income groups, A and B, are able to pay \$5,000 and \$20,000 for land, respectively. The affordable density for each group will vary by distance to the center as represented in Figure 9, where the left axis represents densities and the right axis represents land prices. It should be noted that there is a minimum theoretical density affordable for each group at any distance from the center. This does not mean that each group's bid price will necessary follow the density curve for each group. The two density curves for each group represent only the density, and thus, the area of land that can be purchased for the amount of money each group is willing to pay for land. For instance, Group A can afford to live 3 kilometers from the city center at a minimum density of about 220 p/ha ($45m^2$/person), while at the same distance Group B can afford a minimum density of 60 p/ha ($166m^2$/person).

Figure 8. Land Prices by Distance to the City Center in a Monocentric City

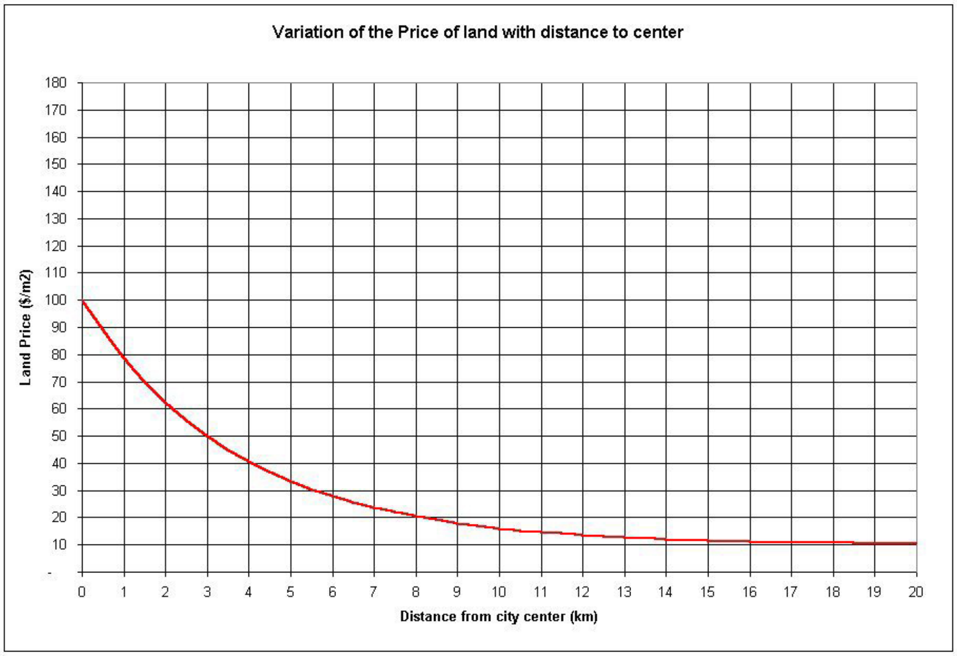
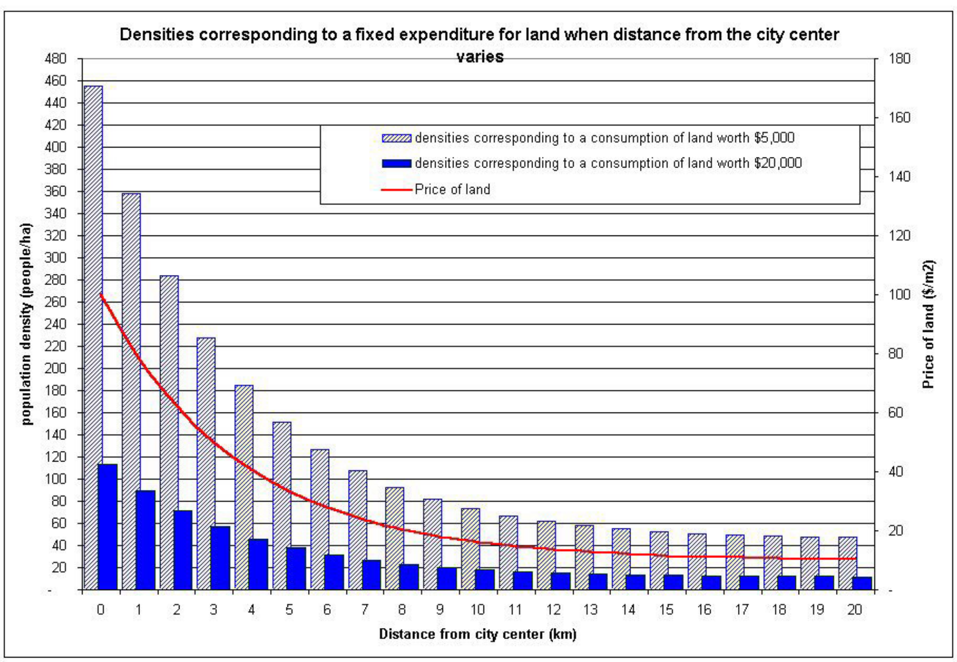


Figure 9. Affordable Density for Two Income Groups by Distance from the City Center



Let us now assume that a well-intentioned urban planner draws a zoning plan covering the entire city, containing—as zoning plans always do—restrictions on minimum plot size, floor area ratio, set backs, etc. The cumulative effect of all these restrictions will result in a de facto upper limit on densities set within the boundaries of each zone. The upper limit on density by distance to the center resulting from the zoning plan is represented by the red, dotted line in Figure 10. It shows that the result of the zoning plan is to exclude Group A from most areas of the city, except between a distance of 3 and 4 kilometers and beyond 14 kilometers from the center. In this particular case, Group A would be practically relegated to the periphery of the city. An alternative for Group A would be to bypass the effect of regulations by switching to the informal sector, but in doing so, Group A will lose part of its property rights.

It must be noted that, in this example, Group B will likely be a strong supporter of the new zoning as it does not affect its residential mobility and prevents Group A from overcrowding the schools and infrastructure of the most desirable part of the city.

The example above is theoretical, but it reflects the realities in many cities of the world. In Brasilia, for instance—one of the most carefully planned cities in the world—the majority of poor households are located in the far periphery while higher income groups are clustered around the center (Figure 11).

Subsidized low-income housing projects usually reduce the residential mobility of poor households. The designers of low-income housing projects must compromise between distance from city center and plot area in advance of project construction. Low-income households, possibly preferring different trade-offs, are therefore tied to a location by the subsidy which goes with the project. Providing housing vouchers to poor households is the only way to direct housing subsidies without reducing the residential mobility of subsidy beneficiaries.

Should municipalities attempt to change their cities' spatial structure?

Because of the different advantages inherent to some spatial structures, as discussed above, logic seemingly recommends changing an existing urban shape in order to meet specific objectives—for instance, a reduction in motorized trips. In doing so, the following two points need to be taken into account: first, there is no optimal spatial structure among the many types of spatial organization; however, a positive density gradient and dispersed incontiguous urbanization are clearly more costly to operate,

Figure 10. Zoning Restrictions and Affordable Densities

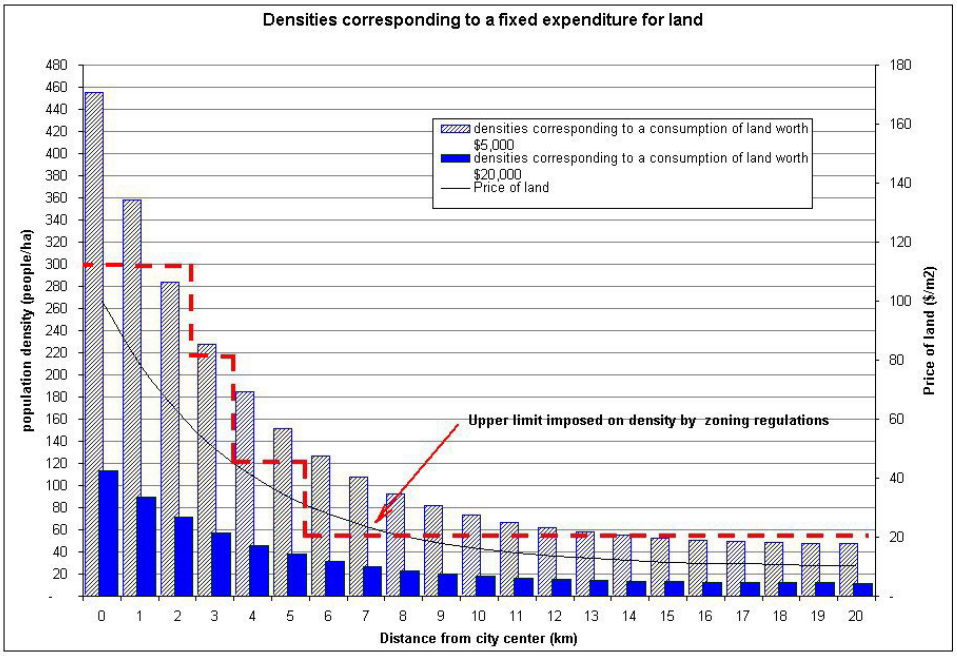
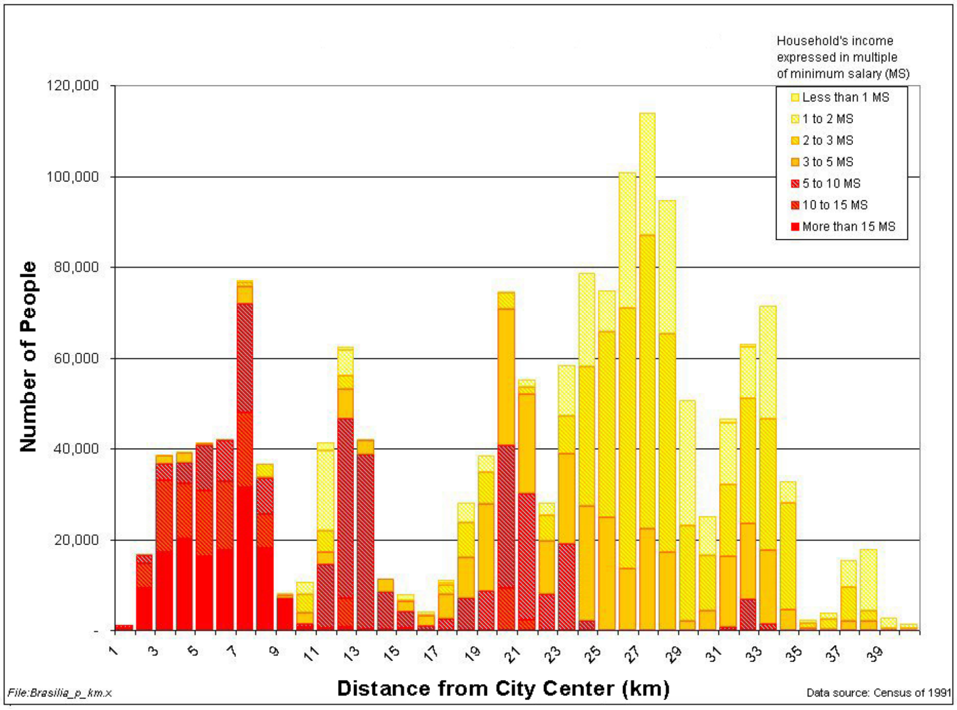


Figure 11. Brasilia: Spatial Distribution of Population per Income Group



have many negative environmental side effects, and should be avoided. Second, urban spatial structures are very resilient and are path dependent—i.e., it is easier to decrease density than to increase it, and it is easier for a monocentric city to become polycentric than the opposite.

Urban planners can influence city shapes only indirectly. In the long run, market forces are building cities (with few unfortunate exceptions like Brasilia). Market forces, however, respond to constraints imposed by regulations and taxation and to opportunities provided by the primary infrastructure network built by the state. Planners, therefore, have only three tools at their disposal to influence urban spatial structure: land use regulations, infrastructure investments, and taxation. Figure 12 shows a schematic view of the interaction of markets and government action in shaping urban structures.

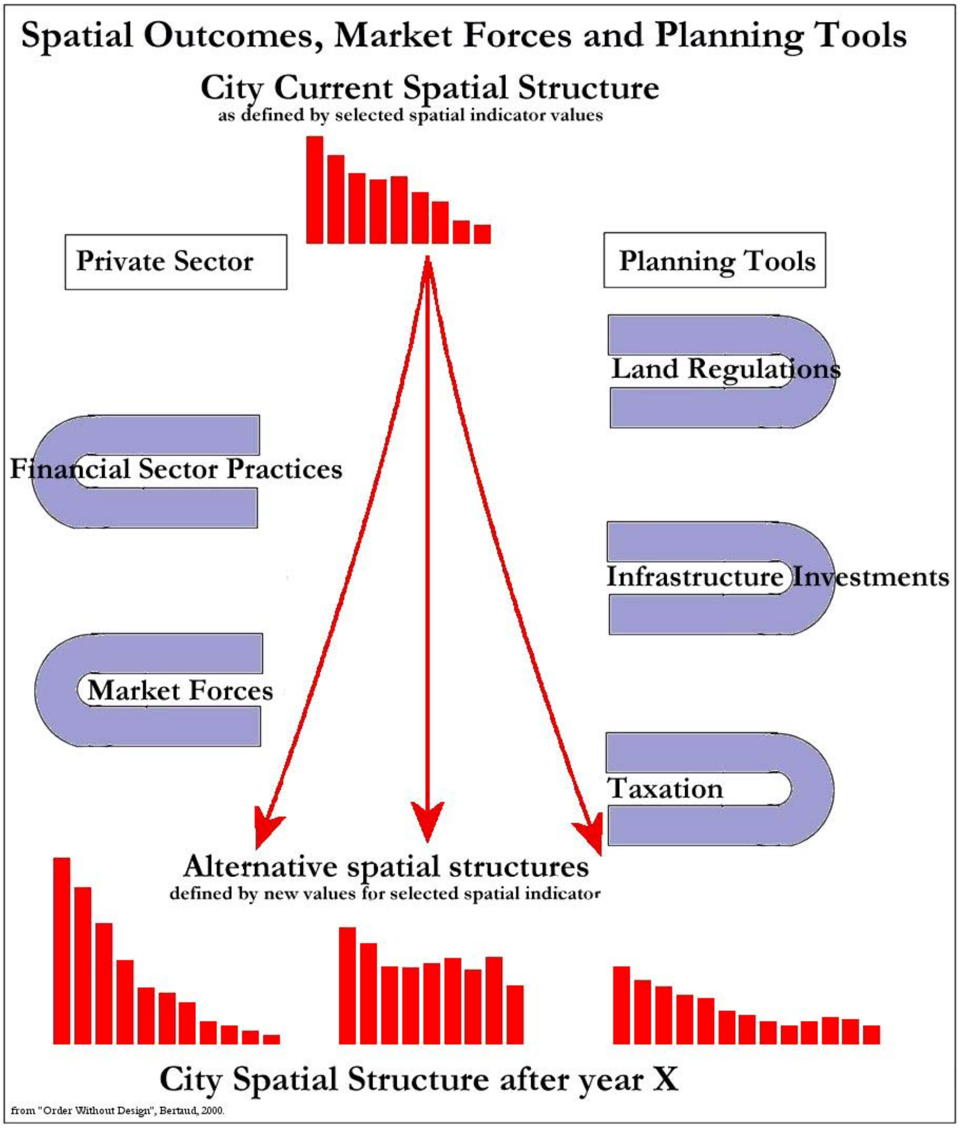
To influence the evolution of a city's spatial structure, these three tools should be carefully coordinated and internally consistent in order to realize a common spatial objective. This consistency is very rare as regulations, infrastructure investments and taxes are often designed at different levels of government and for very different purposes which have nothing to do with a city's spatial structure.

In summary, urban planners should monitor urban structures to be aware of spatial trends and to know the limitations imposed by the current structure on policy options. In some cases, it will be possible to influence spatial trends in a limited way. A more ambitious outcome would require a strong, long-term and continuous political determination combined with a dedicated and well-integrated team of urban planners, municipal and transport engineers, and financial planners. These conditions are seldom met long enough to influence city shape. Curitiba, however, is an example of a city where these conditions were met and where, without doubt, the current city structure is the result of a concerted long range effort. It is, however, not clear whether the resulting spatial structure has resulted in a welfare increase for the majority of the inhabitants of Curitiba, compared to what it would have been had the structure been more market driven.

Is there a global trend in the evolution of urban spatial structures?

Because the development of urban structure is path dependent, dominantly monocentric cities tend to become less monocentric. And because income and mobility are increasing in most large cities of the world, densities in central areas tend to become lower over time. However, by no means are all cities now trending toward the low density, extremely polycentric model.

Figure 12. Interaction Between Market Forces and Government Action



As cities become larger, the CBD itself also becomes larger. However, by becoming larger, the CBD loses the accessibility that made it attractive in the first place. Therefore, it is inevitable that subcenters will emerge as a city becomes larger and the degree of monocentricity decreases with size. However, some very large cities—like New York, London, Buenos Aires and Shanghai—retain a very strong center which, while containing an ever-dwindling ratio of total jobs, remain a very strong attractor for prestige retail, entertainment and culture. By contrast,

some very successful cities manage to grow without any prestigious center. Atlanta and Phoenix are good examples of these types of cities.

The future likely will bring a polarization between the two types of cities, although both types will be polycentric in terms of job distribution. The first type will retain a strong, prestigious center with many amenities, surrounded by a high-density residential area inhabited by mostly high- and middle-income households. The second type will be a pure labor market with no centrally located amenities; jobs and any amenities provided will be evenly distributed throughout the metropolitan area, with no prestigious center.

References

Alonso, William. 1964. *Location and Land Use*. Cambridge, MA: Harvard University Press.

Bertaud, Alain. 1999. *Cracow in the Twenty-First Century: Princes or Merchants? A City's Structure under the Conflicting Influences of Land Markets, Zoning Regulations and a Socialist Past*. ECSIN Working Paper #8. Washington, DC: The World Bank.

Bertaud, Alain, and Bertrand Renaud. 1997. Socialist Cities Without Land Markets. *Journal of Urban Economics* 41:137–51.

Bertaud, Alain, and Stephen Malpezzi. 2001 (September). Measuring the Costs and Benefits of Urban Land Use Regulation: A Simple Model with an Application to Malaysia. *Journal of Housing Economics* 10(3):393–418.

Bertaud, Alain and Stephen Malpezzi. 1999. *The Spatial Distribution of Population in 35 World Cities: The Role of Markets, Planning and Topography.* Madison, WI: Center for Urban Land & Economic Research, University of Wisconsin.

Bollinger, Christopher R., and Keith R. Ihlanfeldt. 1999. *Intrametropolitan Locational Patterns of People and Jobs: Which Government Interventions Make a Difference.* Tallahassee, FL: Department of Economics, Florida State University.

Bruekner, Jan K. 2001. Urban Sprawl: Lessons from Urban Economics. In *Brookings-Wharton Papers on Urban Affairs 2001*, eds. Willam G. Gale & Janet Rothenberg Pack. Washington, DC: The Brookings Institution.

Carlino, Gerald A. 1979. Increasing Returns to Scale in Metropolitan Manufacturing. *Journal of Regional Science* 19:343–351.

Cervero, Robert. 1989. Jobs-Housing Balance and Regional Mobility. *Journal of the American Planning Association* 55:136–150.

Clark, Colin. 1951. Urban Population Densities. *Journal of the Royal Statistical Society* 114:375–386.

Epple, Dennis; Thomas Romer; and Radu Filimon. 1988. Community Development with Endogenous Land Use Controls. *Journal of Public Economics* 35:133–162.

Fischel, William A. 1995. *Regulatory Takings: Law, Economics and Politics*. Cambridge, MA: Harvard University Press.

Goldner, William. 1955. Spatial and Locational Aspects of Metropolitan Labour Markets. *American Economic Review* 45:111–128.

Hall, Peter. 1988. *Cities of Tomorrow*. pp. 210–219. Oxford: Blackwell Publishing.

Hamilton, Bruce W. 1978. Zoning and the Exercise of Monopoly Power. *Journal of Urban Economics* 5:116–130.

Holtzclaw, John. 1994 (June). *Using Residential Patterns and Transit To Decrease Auto Dependence and Costs.* New York: Natural Resources Defense Council.

Ihlandfeldt, Keith R. 1997. Information on the Spatial Distribution of Job Opportunities within Metropolitan Areas. *Journal of Urban Economics* 41:218–242.

Malpezzi, Stephen, and Yeuh-Chuan Kung. 1997 (January). *The Flight to the Suburbs, Revisited: The Intraurban Distribution of Population, Employment, and Housing Prices*. Paper presented to the American Real Estate and Urban Economics Association, New Orleans.

Mills, Edwin. S. and Byong-Nak Song. 1979. *Urbanization and Urban Problems*. Cambridge, MA: Harvard University Press.

Muth, Richard F. 1969. *Cities and Housing*. Chicago: University of Chicago Press.

Prud'homme, Rémy. 1996. Managing Megacities (La gestión de las megalópolis). *Le Courrier du Centre National de la Recherche Scientifique* 82:174–176. Paris: Centre National de la Recherche Scientifique.

Sveikauskas, Leo. 1975. The Productivity of Cities. *Quarterly Journal of Economics* 89:393–413.