URBAN DESIGN *International* (2002) **7**, 153–179

www.palgrave-journals.co.uk/udi

A theory of the city as object: or, how spatial laws mediate the social construction of urban space

Bill Hillier*

Space Syntax Laboratory, The Bartlett School of Graduate Studies, University College London, Gower Street, London WC1E 6BT, UK

A series of recent papers (Hillier *et al*, 1993; Hillier, 1996b, 2000) have outlined a generic process by which spatial configurations, through their effect on movement, first shape, and then are shaped by, land-use patterns and densities. The aim of this paper is to make the spatial dimension of this process more precise. The paper begins by examining a large number of axial maps, and finds that although there are strong cultural variations in different regions of the world, there are also powerful invariants. The problem is to understand how both cultural variations and invariants can arise from the spatial processes that generate cities. The answer proposed is that socio-cultural factors generate the differences by imposing a certain local geometry on the local construction of settlement space, while micro-economic factors, coming more and more into play as the settlement expands, generate the invariants.
URBAN DESIGN *International* (2002) **7,** 153–179. doi:10.1057/palgrave.udi.9000082

Keywords: society; city design; axial maps; formal typologies; cultural invariants and variations

Movement: the strong force

The 'urban grid', in the sense used in this paper, is the pattern of public space linking the buildings of a settlement, regardless of its degree of geometric regularity. The 'structure' of a grid is the pattern brought to light by expressing the grid as an axial map[1] and analysing it configurationally. A series of recent papers have proposed a strong role for urban grids in creating the living city. The argument centres around the relation between the urban grid and movement. In 'Natural movement' (Hillier *et al*, 1993), it was shown that the structure of the urban grid has independent and systematic effects on movement patterns, which could be captured by 'integration' analysis of the axial map[2]. In 'Cities as movement economies' (Hillier, 1996b) it was shown that natural movement – and so ultimately the urban grid itself – impacted on land-use patterns by attracting movement-seeking uses such as retail to locations with high natural movement, and sending non-movement-seeking uses such as residence to low natural movement locations. The attracted uses then attracted more movement to the high movement locations, and this in turn attracted further uses, creating a spiral of multiplier effects and resulting in an urban pattern of dense mixed use areas set against a background of more homogeneous, mainly residential development. In 'Centrality as a process' (Hillier, 2000), it was then shown that these processes not only responded to well-defined configurational properties of the urban grid, but also initiated changes in it by adapting the 'local grid conditions' in the

*Correspondence: Space Syntax Laboratory, The Bartlett School of Graduate Studies, 1-19 Torrington Place, University College London, London WC1E 6BT, UK. Tel: +44 (0) 171 391 1739, Fax: +44 (0) 171 813 4363,
E-mail: b.hillier@ucl.ac.uk

[1]An 'axial map' is the least set of longest lines of direct movement that pass through all the public space of a settlement and make all connections.

[2]The 1993 paper dealt only with 'global' or radius-*n* analysis, but a series of studies since then have shown that 'local' or radius-3 integration is normally a better predictor of pedestrian movement.

mixed movement areas in the direction of greater local intensification and 'metric integration' through smaller scale blocks and more trip-efficient, permeable structures.

Taken together, the three papers describe aspects of a generic mechanism through which human economic and social activity puts its imprint on the spatial form of the city. The papers do not deal with the patterns of activity themselves, but the theory seems to work because, regardless of the nature of activities, their relation to and impact on the urban grid is largely through the way they impact on and are impacted on by movement. Movement emerges as the 'strong force' that holds the whole urban system together, with the fundamental pattern of movement generated by the urban grid itself. The urban grid therefore emerges as a core urban element which, in spite of its static nature, strongly influences the long-term dynamics of the whole urban system. In the light of these results, we can reconceptualise the urban grid as a system of *configurational inequalities* – that is, the differences in integration values in the lines that make up the axial map – which generates a system of *attractional inequalities* – that is, the different loadings of the lines with built form densities and land-use mixes – and note that, in the last analysis, *configuration generates attraction*.

Space-creating mechanisms

The three papers cited describe a process that goes from the spatial configuration of the urban grid to the living city. But what about the grid itself? Is this arbitrary? Would any grid configuration set off the process? The aim of this paper is to try to answer this question. It will be argued that urban grid configurations are far from arbitrary, but in fact are themselves the outcomes of space-creating mechanisms no less generic than the space-to-function mechanisms described in the three cited papers. The argument runs as follows. If we examine a large number of axial maps, we find well-defined invariants as well as obvious differences. What process, we must ask, can produce both. The answer proposed is that the invariants arise from a combination of two things. First, in spite of all their variability, there are certain invariants in the social forces – or more precisely in the relations between social forces – that drive the process of settlement aggregation. Second, there are autonomous spatial laws governing the effects on spatial configuration of the placing of objects such as buildings in space, and these constitute a framework of laws within which the aggregative processes that create settlements take place. The social forces working through the spatial laws create both the differences and the invariants in settlement forms. The link between the two is again movement, but whereas the 'space-to-function' mechanism was driven by the effect of spatial configuration on movement, the space creating mechanism is driven by the influence of movement on space, and so can be considered a 'function-to-space' mechanism.

The concept of spatial 'laws' is critical to this argument, so we must explain what this means. Spatial 'laws', in the sense the term is used here, does not refer to universal human behaviours of the kind claimed, for example, for the theory of 'human territoriality' (as reviewed in Vischer-Skaburskis, 1974), but to 'if–then' laws that say that if we place an object here or there within a spatial system then certain predictable consequences follow for the ambient spatial configuration. Such effects are quite independent of human will or intention, but can be used by human beings to achieve spatial and indeed social effects. Human beings are bound by these laws in the sense that they form a system of possibilities and limits within which they evolve their spatial strategies. However, human agents decide independently what their strategies should be. Like language, the laws are then at once a constraining framework and a system of possibilities to be exploited by individuals.

In fact, it seems likely that human beings already intuitively 'know' these laws (although they cannot make them explicit), and can exploit them as agents to create social effects through spatial behaviours at a very young age. Consider the following true story. A group of people are sitting in armchairs in my daughter's flat. My two-year-old grandson Freddie comes into the room with two balloons attached to weights by two pieces of string about two and a half feet long, so that the balloons are at about head height for the sitting people. Looking mischievous, he places the balloons in the centre of the space defined by the armchairs. After a minute or two, thinking Freddie has lost interest, one of the adults moves the balloons from the centre of the space to the edge. Freddie, looking even more mischievous, walks over to the balloons and places them back in the centre of the room. Everyone understands intuitively what is going on, including you. But what is actually happening?

The answer is that by placing an object in the centre of a space we create more obstruction to lines of sight and potential movement than if we place it at the edge. This is the principle of 'centrality' set out in the 'theory of partitioning' in Chapter 8 of 'Space is the Machine' (Hillier, 1996a, b). If we place a partition midway on a line, it creates more – and more evenly distributed – gain (added distance in summing shortest trips from all points to all others) in the universal distance (the sum of distances from each point to all others) than if we place it peripherally (in which case the depth gain is more unevenly distributed, but is overall less). Because this must apply to lines in all directions, it follows that it will also work for objects placed in space. An object placed centrally in a space will increase universal distance and interrupt intervisibility more than one placed at the edge. Now it is clear that Freddie not only 'knows' this in the sense that he can make use of this knowledge in behaviour, but it is also clear that he can use this – surely 'theoretical' – knowledge of space to achieve social ends, namely drawing attention to himself and away from the adults engaged in conversation. It is also of course clear that we 'know' this about space in the same way as Freddie, but it is also clear as professionals that it is unlikely that we were taught this vital principle of space in architecture school or in maths class.

What is proposed here is that spatial laws, driven by social forces, account for exactly and only the spatial invariants of cities[3]. The form of the paper will be to: examine axial maps and develop an account of their invariants as well as their differences; outline and demonstrate the spatial laws in question; apply these to what will be called the 'basic generative process' by which urban-type spatial systems arise; and develop a theory of how the impact of the spatial laws on evolving settlements is driven by two kinds of social forces, which can be broadly termed the socio-cultural and the micro-economic. It is proposed that culture is a variable and puts its imprint mainly on the local texturing of space, generating its characteristic differences, whereas micro-economics is a constant and puts its imprint mainly on the emerging global structure of the settlement in a more or less invariant way. The reason one works locally and the other globally is due to the ways in which each uses the same spatial laws to generate or restrain potential movement in the system.

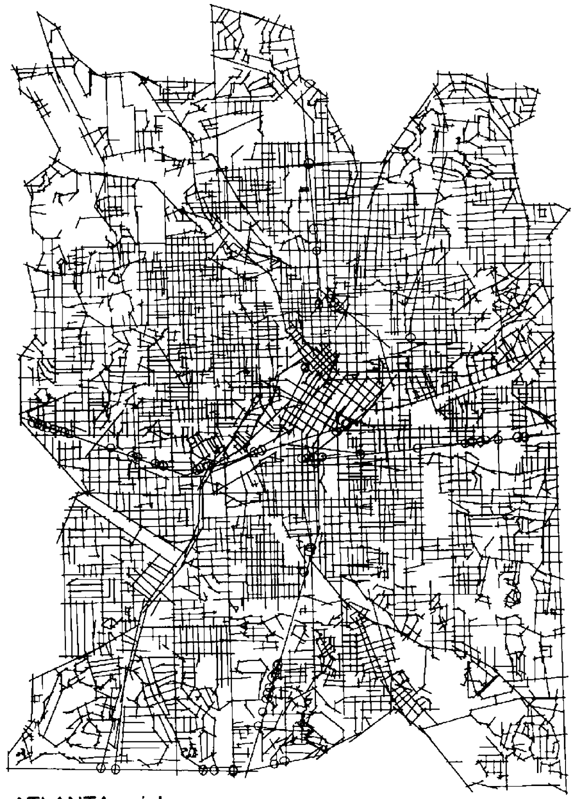
This is why we find in axial maps both differences in local texture, and invariants in the global patterning. The combination of the spatial laws and the dual processes explains why axial maps read as a set of similarities and differences. The paper concludes with a discussion of the relation between socio-economic and spatial laws, suggesting that although the creation of the space of the city is driven by socioeconomic processes it is not shaped exclusively by them. Equally fundamental in shaping city space are autonomous spatial laws that generate more or less equifinal outcomes from varying processes[4].

Differences and invariants in axial maps

First, let us consider some axial maps. By far the most obvious differences between them are geometrical. On reflection, that is all they could be. Axial maps are no more than sets of lines of different lengths with different angles of intersection and different degrees and kinds of intersection (for example, a line can either pass through another or stop on it). Axial maps from different parts of the world tend to differ in all these properties. Figures 1–4 show four fairly characteristic axial maps from different parts of the world arranged from the most to the least 'geometric': Atlanta (USA), The Hague (Holland), Manchester (UK) and Hamedan (Iran). It is easy to see that the impression of 'more to less geometric' arises because the axial maps differ substantially on the basic properties of axial maps. Each has its own distinctive range of line lengths and angles of the incidence, and its distinctive intersection characteristics. For

[3]In this sense, the argument is still within the spirit of the theoretical framework set out by Martin and March in *Urban Space and Structures* in 1972.

[4]The question 'What about planned towns?' may of course be raised here. However, in the great majority of cases the planned element is only the first stage of an urban growth process that then will be subject to the same lawful influences as cities which have grown through a distributed process.



ATLANTA axial map

Figure 1.

Figure 2.

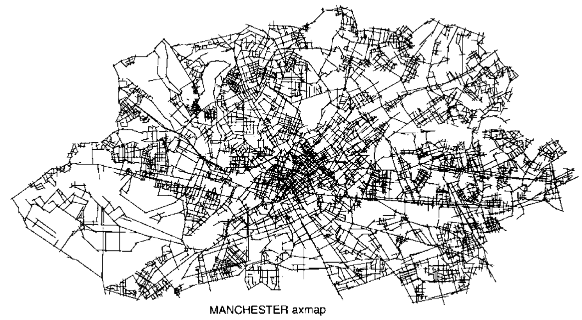


Figure 3.

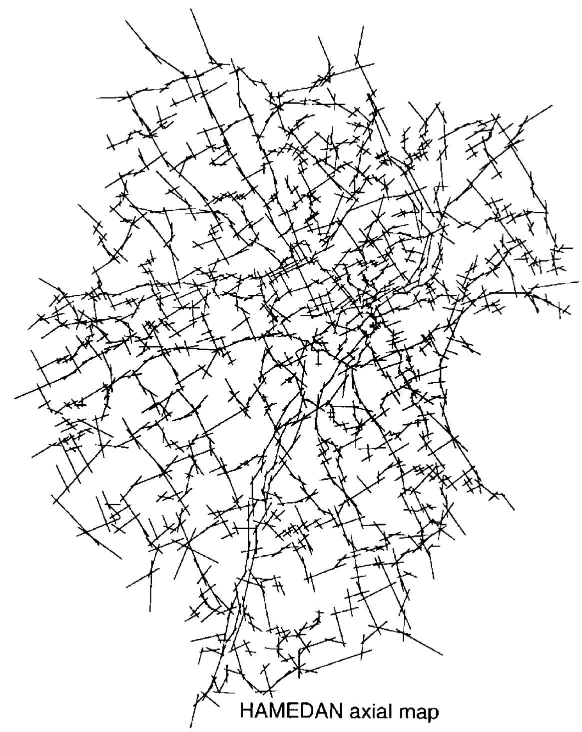


Figure 4.

example, if we use the patterns of intersection, we find that in Atlanta, the tendency for lines to pass through each other (rather than to end on other lines) is very marked at all levels. In The Hague, this is found locally but much less so at the global level. In Manchester, this is hardly found globally, and what there is locally is much more broken up than in The Hague. In Hamedan, it hardly exists either at the global or local levels, except in the central public areas of the town.

Differences in the range of line lengths and angles of incidence seem to follow the intersection differences. Atlanta has a number of very long lines approximating the radius of the system, and long lines can be found in most parts. At the same time large areas of the grid maintain a strict right-angle intersection with a north–south orientation,

Table 1

	Cities	Avg. Lines	Conn	Loc Int	Glob Int	Intel
Usa	12	5420	5.835	2.956	1.610	0.559
euro	15	5030	4.609	2.254	0.918	0.266
uk	13	4440	3.713	2.148	0.720	0.232
arab	18	840	2.975	1.619	0.650	0.160

Conn – connectivity; Loc Int – local integration; Glob Int – global integration; Intel – intelligibility.

although with a striking offset grid in the historic centre. In The Hague, the longest lines tend to be less than the radius of the system, and in general long lines are peripheral to discrete local groups of lines. A less strong right-angle discipline is maintained, there is greater variety in the orientation, and long lines, especially radials, tend to intersect to others at very obtuse angles. In Manchester, the long lines are nearly all radials well below the radius of the system. The tendency for the long radials to have near-straight continuations is even stronger, and the local right-angle discipline is even looser. In Hamedan, the longest lines are only a fraction of the radius of the system and tend to be found towards the periphery of the system. Even so there is a clear radial structure formed by lines of the second length rank, and intersecting with greater angular change than in the other cases. Locally, we find a whole range of angles of incidence including near-right-angle connections, but in most cases one line tends to stop on another.

These geometric differences are also consistently reflected in syntactic differences.

Table 1 shows the syntactic average for 58 cities taken from four parts of the world. Each regional group of cities, in spite of differences within the subsamples, has its own characteristic set of syntactic parameters.

What is the reason for these geometrical and syntactic differences? Why should lines in Iranian cities be, on average, markedly shorter than lines in, say, English cities, or why should European cities have a degree of geometric organisation somewhere between UK and American cities, or Arab cities be less intelligible than European cities. On the face of it, the differences seem to be expressions of what we might call 'spatial culture'. For example, in cities in the Arab world, the spectrum between public and private spaces is often quite different from that in European cities. In historic European cities, we find that local areas are for the most part easily permeable to strangers, with public spaces in locally central areas easily accessible by strong lines from the edge of the area. At the same time, fronts of dwellings are strongly developed as facades and interface directly with the street both in terms of visibility and movement. In many Arab cities, strangers tend to be guided much more to certain public areas in the town, and access to local areas is rendered much more forbidding by the more complex axial structure. At the same time, dwelling facades are much less developed, and the interface with the street tends to be much less direct both for visibility and for movement. The differences in the geometry of the axial maps seem to be a natural expression of these differences. Even in the case of American cities, where one of the main factors in creating the more uniform American grid is thought to be the need to parcel up land as quickly and easily as possible to facilitate economic development, we note that the grid was prior to economic development and should therefore be seen as a 'spatial cultural' decision to create and use space in a certain way.

However, in spite of these differences, there are also powerful invariants in axial maps that seem to go across cultures and even across scales of settlement. One of the most striking is the statistical distribution of line lengths. Although we find great variations in the average and range of line lengths, we invariably find:

– that the axial maps of cities are made up of a small number of long lines and a large number of short lines;
– that this becomes more the case as cities become larger; and
– that in general the distribution of line lengths in cities approximates a logarithmic distribution.

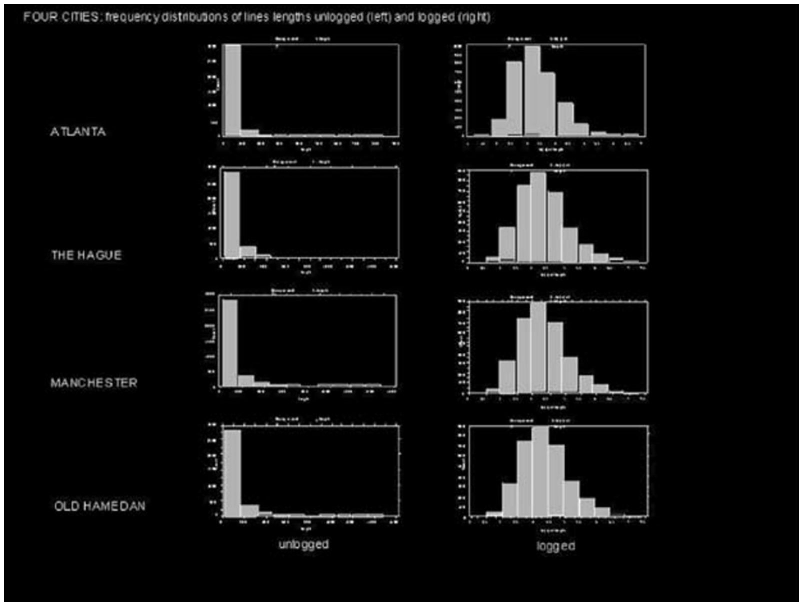


Figure 5.

Figure 5 shows the four cities of Figures 1–4 with the distribution of line lengths on the left and the logged distribution on the right[5].

In practical terms, this means that if, for example, we divide the range of line lengths into 10, we find that in Atlanta 92.7% of lines are in the decile of shortest lines and only 2% in the eight longest. In The Hague, the figures are 84.8 and 5%, and in Manchester 85.9% and less than 3%. In the much smaller case of Hamedan, we find that 90% of the lines are in the four shortest deciles and only 2% in the five longest. Looking more widely, we find that in London (15,919 lines) 93.3% of lines are in the shortest decile and less than 1% in the top eight deciles. In Amsterdam (7996 lines) – on the face of it a more griddy city – the figure is 95.8% in the shortest decile and again less than 1% in the top eight. In Santiago (29,808 lines), an even more grid-like structure, the figure is 94.7%, again with 1% in the top eight, while in Chicago (30,469 lines), a city we think of as wholly grid-like, the figure is 97.6% with only 0.6% in the top eight. It is not always quite so high in large cities. In Athens (23,519 lines), for example, the figure is 86%, with 2.3% in the top eight. However, even in the strange pre-Columbian city of Teotihuacan the figure is 85%.

If we look at a smaller system, we find the same tendency, although less marked. In Venice (2556 lines), for example, the figure is 76.3% with 4% in the top eight; in Shiraz (1971 lines) in Iran, where lines are on average shorter than in Western cities, 71.7% are in the shortest decile and 8.3% in the top eight. In the English cities of Nottingham, Bristol and York, the figures for the shortest decile are 78, 63 and 55%, respectively. Even in much smaller systems, we can find a strong tendency in this direction. If we take Old Paranoa, the informal settlement built by the workers who constructed the dam in Brasilia (de Holanda, 1977), we find that 32% of the lines are in the shortest decile and 68% in the shortest two. In the southern French town of Apt, 41% are in the shortest decile and 59% in the shortest two, while in Serowe (a self-generated settlement in southern Africa) 32% are in the shortest decile and 68% in the shortest two. Even in a small area within London we find 24% in the shortest decile and 53% in the shortest two.

As settlements grow, then, the proportion of lines that are long relative to the mean for the settlement becomes smaller but the lines

[5]In some cases, such as Chicago and Amsterdam, we find a loglognormal distribution. However, the difference between a log and a loglog distribution is much greater than that between an unlogged and logged distribution, so these differences are not pursued here.

*Note: Only selected images have been printed along with the text here.

themselves get longer. This seems to be invariant across all cultures in spite of the strong geometric differences we have noted. A plot of the log of the number of lines against the proportion in the shortest decile for 20 settlements from small to very large, showing an r^2 of 0.802 ($p = 0.0001$)[5]. This also applies to different-sized chunks of the same city. If we plot the percentage of lines in the shortest decile against the number of axial lines for four different-sized cutouts from the London axial map, showing an r^2 of 0.923, $p = 0.0391$. However, even the smallest cutout – the City of London with only 565 lines (as opposed to 15919 for the largest system) has 70% of lines in the shortest decile, and approximates a logarithmic distribution.

Why then are line lengths distributed in this way and, in particular, what is the role of the small number of long lines? A useful clue comes from looking at their spatial distribution. If we take the lines in the longest quintile of the range and make them the darkest lines in the axial map, we find a marked tendency for the longest lines to be centre to edge lines starting at some distance from the original centre. Figure 6 shows the pattern for London, and Figure 7 for Athens. The second rank of lines, however, shows a different pattern in each case. In London, the second-rank lines form a continuous and relatively dense network penetrating most parts of the grid. In Athens, the second-rank lines pick out discrete grid-like areas with relatively poor connections between them. If we were to look at, say, Baltimore, the second-rank lines tend to be linked directly to the first rank of lines, forming a tree-like distribution in

Figure 6.

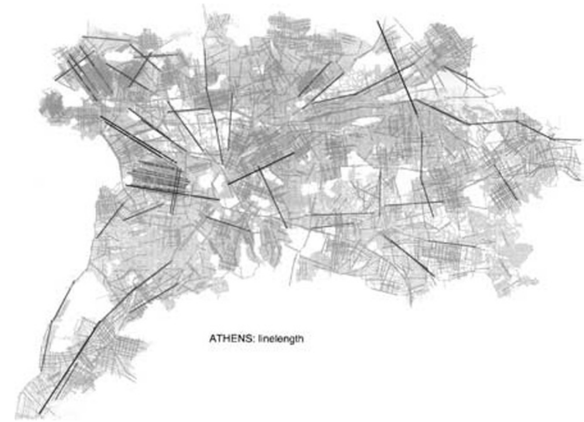


Figure 7.

the system as a whole. These patterns suggest that the first rank of lines reflects generic properties of city growth while the second rank indicates differences in the relation of global to local.

This hint of global invariants and local differences is reinforced if we look at the syntactic analysis of the axial maps. If we take the four cities shown in Figures 1–4 and analyse them for radius-n integration (for example, Figure 8), we find in each case that in spite of the geometric differences a certain kind of structure is adumbrated: each city has an 'integration core' – the patterns formed by the darker – which links a grid-like pattern of lines at the heart of the city almost to the edge in all directions either by way of quasi-radial lines or extended orthogonal lines, in some cases reaching the edge line but in others falling short. Within the interstices formed by this overall lighter areas are found, often with a darkish line as a local focus. In other words, in spite of the geometric differences, each city has, when seen as a system of configurational inequalities, a certain similarity of structure. This is the pattern we call the 'deformed wheel': a hub, spokes in all main directions, sometimes a partial rim of major lines, with less integrated, usually more residential, areas in the interstice forms by the wheel. This generic pattern was first identified as a deep structure common to many small towns, seeming to occur in spite of topographic differences (Hillier, 1990). It was also found as a local area structure in London, where named areas such as Soho or Barnsbury typically took an area deformed wheel form with the London 'supergrid' (the main

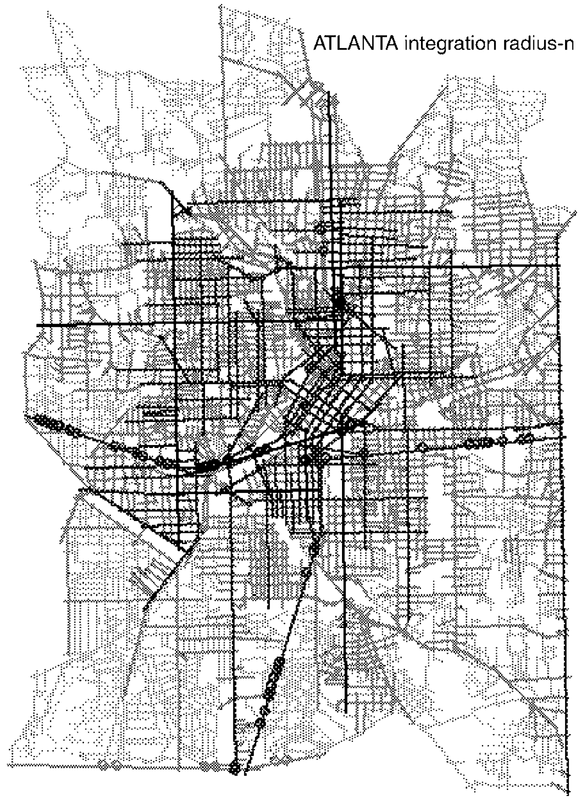


Figure 8.

radials and their lateral links) forming the rim of the wheel.

As a global pattern, the deformed wheel holds up remarkably well in larger cities for example, in London, Athens and Baltimore. The pattern is even found in very different kinds of cities. If we look at Venice (Figure 9) without the canals, for example, we find that in spite of its very idiosyncratic history – having grown together from several islands rather than from a single origin – we still find a very marked 'deformed wheel' pattern, even though the wheel is much less easily recognisable than in most cases. Or looking at Tokyo, which is by far the largest system ever analysed, we find a remarkable and even more complex version of the wheel pattern with several layers of rim which, with the sinuous radials, produce a quasi-grid covers a large part of the system. Even the strange pre-columbian city of Teotihuacan shows, at least a partial realisation of the deformed wheel pattern. Again this near invariant of cities is found in spite of the substantial differences in syntactic values that were shown in Table 1.

Figure 9.

In addition to these space invariants, we also find that if we look at settlements in terms of the size and shape of blocks, then we find if not invariants then at least a set of pervasive tendencies, once again set against a background of substantial geometric differences by region (which we may therefore expect to have a cultural origin of some kind). These can be seen fairly easily in the axial maps of Figures 1–4, and even in the analysed axial maps, but is perhaps easier to see in black on white figure ground maps of a Turkish city analysed by Sema Kubat shown in Figure 10. The most obvious 'near invariant' is an underlying tendency for blocks to be smaller and more convex at or near the centre and larger and less convex towards the edges (Hillier, 2000). However, if we relate block size and shape to the patterns shown by integration analysis of the axial map, we find a subtler pattern. The lines forming the spokes of the deformed wheel tend to be lined with larger than average (for the settlement) blocks for most of their length, but smaller than average blocks in the centre (the hub of the wheel).[6] In contrast, the areas interstitial to the core tend to have block sizes between these two extremes. In other words, the distribution of block sizes seems to reflect the distinction between global and local structure. This is to some extent the case in all the settlements shown so far.

[6]This phenomenon is also found in subcentres. In 'Centrality as a process' (Hillier, 2000), it was argued that wherever movement is convex and circulatory (ie moves around in a locally two-dimensional grid as eg in a shopping centre) rather than linear and oriented (as in moving through an urban grid from an origin to a destination), then metric integration was the key property in understanding both the movement pattern and the type of spatial configuration that tended to emerge under these conditions.

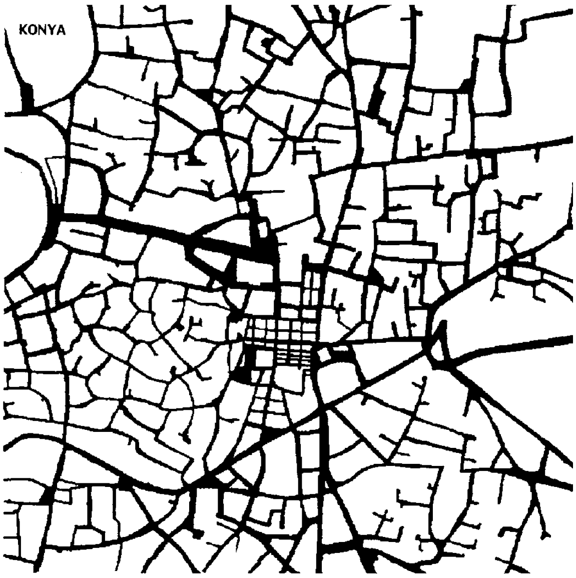


Figure 10.

Figure 11.

Socio-cultural relativities and economic universals

We are faced then with a puzzle. The processes that generate the axial maps and block maps of cities seem at the same time to produce variants, in the form of systematic differences in settlement geometry and syntax from one region to another, and also invariants. What kind of process can produce both? It seems highly unlikely that these dual patterns are in any sense 'designed in', although of course they may be in some cases. However, the fact that most settlements evolve over long periods compels us to the view that the patterns arise from a largely 'distributed' or 'bottom-up' process, that is, from multiple interventions by many agents over time. Even if single agencies are involved, then even so the fact that settlements evolve over such long periods implies that the process of settlement generation must be regarded as an essentially distributed one. What kind of distributed process, then, can produce such dual emergent phenomena?

Let us first note an important difference between variants and invariants: the variants tend to be local and the invariants global. Now consider a case where a city has grown under the influence of at least two different cultures: Nicosia. Figure 11 is an analysed axial map of the historical core of Nicosia in Cyprus (a city sadly now divided). The north-east quarter is a historic Turkish area, the south-east a historic Greek area. The differences in the texture of the grid are marked, with the two areas having quite different geometries and different emergent topologies: the Greek area has longer lines, more lines passing through each other, a different pattern of angle of incidence, and, as a result, much more local and global integration (and a better relation between the two) than the Turkish area. Since these differences reflect typical differences found between systems in Europe and the Islamic world, it is reasonable to regard these as socio-cultural differences in the basic geometry of space. However, when we analyse the area as a whole we find a typical deformed wheel pattern has somehow arisen over and above these geometrical differences, even though the differences between the Greek and Turkish areas show up strongly as differences in the degree of integration.

We thus see what appear to be two processes operating in parallel: one a local process generating differences in local grid patterns and apparently reflecting differences in spatial culture in some way; and the other a global process generating a single overriding structure that seems to reflect a more generic or universal process of some kind. A clue to this comes from the simple fact that the less integrated areas generated by the local process are largely residential, and it would be natural to think of these as the primary distributed loci of socio-cultural identities, it being through domestic space and its environs (including local religious and cultural buildings) that culture is most strongly reproduced through the spatiality of everyday life. A second clue comes from the fact that when we analyse settlements syntactically, it is the micro-economic activity of markets, exchange and trading that is most strongly associated with the 'ntegration core', religious and civic buildings being much more variably located (Loumi, 1988; Karimi, 1998; Hillier, 2000). In this, of course, the integration core of public space also reflects the spatiality of everyday life, but in this case it tends both to the global, because micro-economic activity in its nature will seek to extend rather than confine itself, and also to be culturally nonspecific, in that it is in these activities, and therefore these spaces, that people mix and cultural differences are backgrounded.

This suggests a natural explanation for the dual production of variants and invariants in urban grids. On the one hand, a residential process driven by socio-cultural forces puts its imprint on local space by specifying its geometry and generates a distinctive pattern of local differences, because culture is spatially specific. On the other, a public space process driven by micro-economic activity generates a globalising pattern of space that tends to be everywhere similar because micro-economic activity is a spatial universal. This is the critical difference between the two aspects of the settlement creating process: the socio-cultural component is idiosyncratic and local while the micro-economic component is universal and global. It is this that creates the underlying pattern of differences and invariants that we find everywhere in settlement forms.

This is the key conjecture of this paper: that the processes that generate settlement forms are essentially dual, and through this duality generate the invariant pattern of local differences and global similarities that characterises settlement forms. The question then arises: why should socio-cultural life generate one kind of spatial pattern and micro-economic life another? The answer, it will be proposed, lies in the fact that the relation between micro-economic activity and space, like the relation between culture and space, is largely mediated by movement, but micro-economic economics in a universal and global way, culture in a local and specific way. In what follows we will therefore look at spatial and movement aspects of both socio-cultural and micro-economic processes and how they affect each other as a settlement grows.

The basic generative process

We can begin by noting that there is also a set of low-level invariants, or near-invariants, in urban space, which are so commonplace as to be rarely remarked on, but which are the very foundation of what a settlement is. These are

– that most spaces are linear, defined by the entrances of buildings or groups of buildings on both sides;
– that buildings are clumped together to form discrete islands;
– so that the linear spaces surroundings the islands form intersecting rings and create an overall system of continuous space (a 'street pattern' of some kind); and
– that this is a highly nondendritic configuration, that is a pattern that is everywhere ringy rather than tree-like.

The simplest process for generating spatial configurations with these properties has been familiar since the earliest days of space syntax: the restricted random 'beady ring' process that generates small ring street settlements of a kind found in many parts of the world (Hillier and Hanson, 1984). The process starts with a dyad composed of a cell (representing a notional building) and a piece of open space linked by an entrance so that those inside can come and go into the outside world. These dyads aggregate randomly apart from two restrictions: that each open cell must join full-facewise onto one already in the system (joins of closed cells arise only randomly); and no vertex joins for closed cells are allowed (people do not build corner to corner)[7].

The pattern on the left of Figure 12 is a typical product of such a process. A 'beady-ring'-type pattern is produced on the way, but this is not our main concern here. The overall pattern is that a system of outward facing islands of built forms varying in size and creating more or less linear spaces forming intersecting rings has emerged from the process. No one designed this. It has emerged by a process that finds a pathway of emer-gence by which a global pattern appears

[7]In the version of the process set out in *The Social Logic of Space*, the open space of the dyad was the same size as the built cell. In the version shown here, this has been retained, but the built cells have then been expanded without expanding the open spaces, with the effect that the scaling of open spaces and buildings approximates real systems more closely.

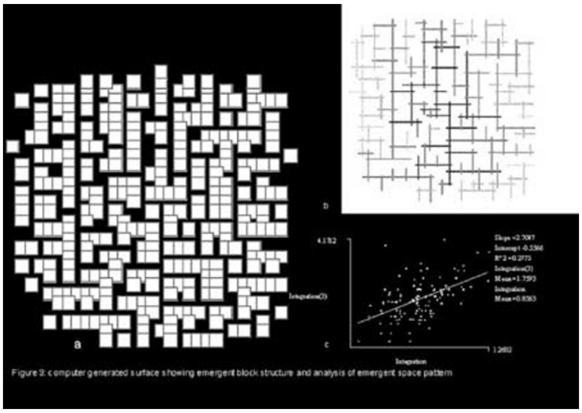


Figure 12.

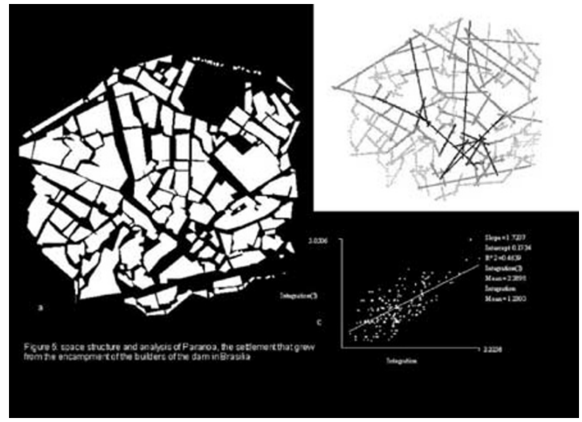


Figure 13.

from the actions of local agents. A key element of the urban system has thus emerged in the form of a continuous system of open space, permitting interaccessibility from each part of the settlement to all others.

The pattern thus has enough of the key topological settlement-like properties (although it lacks their geometric properties – but see below) for us to think of it – by Ockham's razor perhaps – as the 'basic generative process' for spatial patterns of a generically urban kind. But it does not yet look at all like a real settlement. What is missing? It cannot be just the over- regularity due to the fact that the process has been generated on a regular grating. The fault seems to lie mainly in the geometry of both its block structure and its line or axial structure: blocks are insufficiently compact and lines are insufficiently varied in length. So let us look at two real settlements that seem to have grown by something like this process and see what they have in addition. Figure 13 is the old self-generated settlement of Paranoa, developed from the encampment of the workers who built the dam for the lake behind Brasilia (de Holanda, 1997). Figure 14 is the settlement of Serowe in south-west Africa in which the built elements are actually compounds. On the right are radius-*n* integration maps of each, and the 'synergy' scattergram plotting the correlation between local and global integration. On the right of Figure 12 is the same analysis of the computer-generated pattern.

Two points are of particular interest. First, something like the deformed wheel 'integration core' exists in both real cases (and in the case of Paranoa cannot be explained in terms of existing routes in the direction of other settlements, since there were none except to the south). Second, when we look at the 'synergy' scattergrams, we find that the r^2-between local (radius-3) and global (radius-*n*) integration is much better than in the generated case in spite of the fact that it lacks the discipline of an underlying grid. In other words, Paranoa and Serowe both display a relation between local and global structure that needs to be explained.

Experiments with random lines

We can explore these differences further by experimentation. We first construct a more or less random rectilinear grid made up of lines that vary in length only a small amount, on average about half the diameter of the overall 'settlement'. The scattergram gives an r^2 between connectivity and integration of over 0.8[8,9]. We then retain the same mean and range of line length but grow the system to twice its size. Its diameter is now about three times the mean line length. The 'intelligibility' r^2 falls to 0.5. We do the same again, increasing the size of the system until its diameter is about four times the mean line length. The r^2 falls to below 0.3, as in Figure 15.

It is not difficult to work out what is happening. If integration analysis is carried out on a system

[8]Old Paranoa has now been pulled down by the planning authorities and replaced by a much more regular settlement.
[9]The 'intelligibility' correlation between connectivity and global integration is used here rather than the 'synergy' correlation between local and global because the systems are initially too small to respond realistically to local integration analysis. The argument would however hold up for 'synergy' analysis.

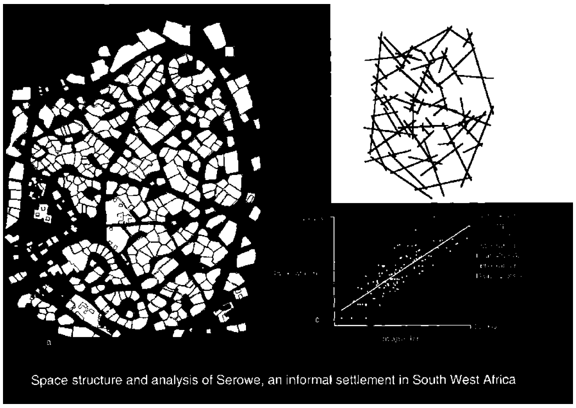


Figure 14.

with uniform elements much smaller than the system itself – say a tessellation of square cells – then integration will focus on the geometric centre and fall off towards the edges. As soon as you specify a system with more or less similar dimensions, in this case similar line lengths, then the same must happen. As the system becomes larger, integration will increasingly concentrate in the centre. The consequences for the 'intelligibility' relation between line connectivity (which is closely related to length) and integration are that relatively longer and therefore better connected lines will be randomly distributed through the system, while integration will be concentrated in the centre. The more this happens, the less the two will correlate and the more the local properties of the system give a poor guide to the global properties – hence unintelligibility.

If, we then take four lines near the centre and extend them to a length of about 0.75 of the diameter of the system, the effect on both the integration core and the scattergram is immediate and dramatic. The core, not surprisingly, begins to go from centre to edge and the scattergram improves from below 0.3 to above 0.6. However, the scatter is highly non-urban, in that the four new lines are quite distinct from the rest of the system. But in Figure 16, an r^2 of .86 is achieved with a pattern of lines that links laterally at the edges as well as from centre to edge: the characteristic deformed wheel structure. In this axial map, 47% of lines are in the shortest decile and a further 29% in the next shortest, almost identical to Paranoa, where the respective figures are 52 and 25%.

This suggests that the essential function of the longer lines against the background of shorter lines is, as we might expect, to give some kind of global structure to the overall pattern, with the local structure fitted into its interstices. However, two further points must be added. First, we also find that the pattern of long to short lines is critical not just to the global structure but also to the relation between the local and global structure. This suggests that the long to short distribution is pervasive at all levels of the settlement and its growth, and therefore needs to be understood as an outcome of a growth process rather than as one of imposition of a global structure. In other words, we need to

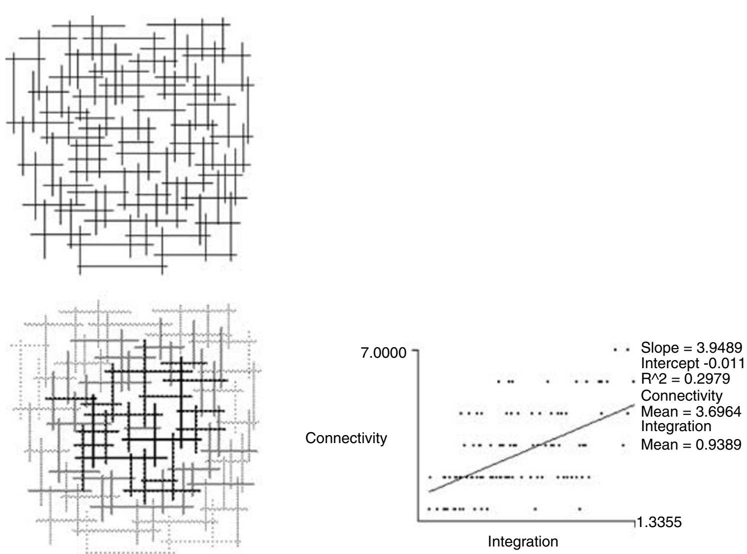


Figure 15.

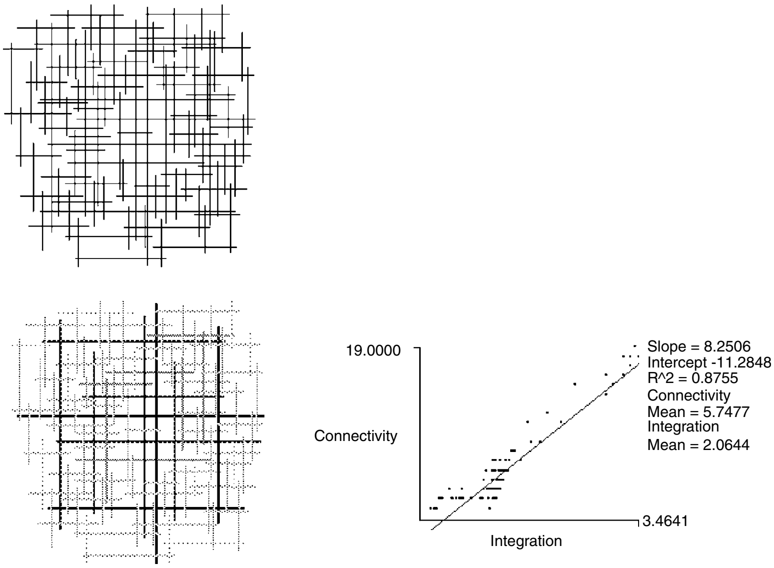


Figure 16.

understand how the required distribution of line lengths can be produced at every stage of an aggregative process of settlement growth.

Regularities in the configurational effects of placing objects

How then can we modify the 'basic generative process' to create these outcomes at every level, so that the growth process will tend to create not only a pervasively lognormal distribution at every level, but at the same time generate an intelligible and synergic system with a deformed wheel-type structure? The answer proposed is that it is here that spatial laws intervene, driven by the dual socio-cultural and micro-economic forces imposing on space their different requirements for potential movement.

The laws in question govern the effects on spatial configuration of the placing objects (such as buildings) in space. The laws initially govern the degree of metric integration in the system measured as the universal distance[10] from each cell in the complex to all others (as opposed to a specific distance that measures distance from one cell to one other). The mean universal distance in a complex is thus isomorphic to the mean length of the trip by shortest paths within the complex. It is through their effect on mean trip lengths that these laws are activated and govern the evolution of the urban object.

The laws are essentially clarifications, simplifications and fuller demonstrations of the 'principles of partitioning' set out in Chapter 8 of *Space is the Machine*. There it was shown that every time a partition is placed in a system, it has a predictable effect on 'universal distance' within that system. In that text, four partitioning principles were proposed for the minimising or maximising of 'depth gain' in a system, depth gain being the increase in universal distance due to the placing of a partition. The principles were: *centrality* – partitioning a line in its centre creates more depth gain than partitioning it eccentrically; *extension* – partitioning a longer line creates more depth gain than partitioning a shorter line; *contiguity* – making partitions contiguous increases depth gain more than making them discrete; and *linearity* – arranging contiguous partitions linearly increases depth gain more than coiling them up, as, for example, in a room.

In what follows it will be proposed that these four principles[11] can be reduced to two laws, one dealing with the relations of spaces and the other

[10]For an account of the idea of 'universal distance' see *Space is the Machine*, Chapter 3. Universal distance is probably the most fundamental concept in space syntax. It can be applied either metrically or topologically, and allows the redefinition of an element in a system as no more than a position from which the rest of the system can be seen, thus 'nearly dissolving the elements'. (see Space as paradigm in the Proceedings of the Brasilia Space syntax Symposium).

[11]On reflection, what were noted in *Space in the Machine* were empirical regularities, since no theoretical account was offered as to why they should be so.

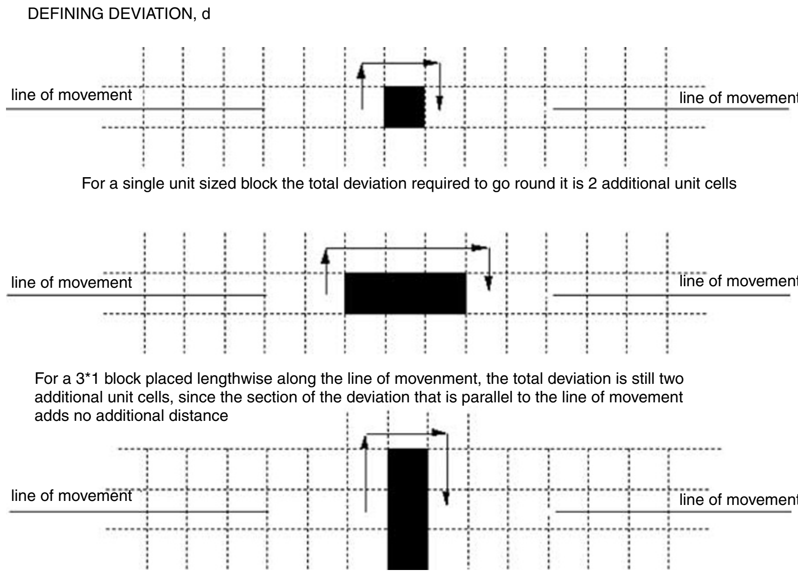


Figure 17.

with the relations of objects. Before we introduce these laws however, we will show how these partitioning regularities can be interpreted for cellular aggregates. The basic notion we work with is that of a pair of cells (or boundaries) forming the two ends of a line, and a third cell which we wish to place between them. The method for calculating the gain in universal distance is as in Figure 17. Consider a line of $n+1$ cells with an object placed somewhere along it leaving n cells in some distribution on the two sides of the cell with at least one cell on each side. A deviation, d, will be the unit distance around the object that must be added to straight line movement to go from any cell to any other on the other side of the object. D will be the sum of d's that are needed to go from all cells to all others, or the total added to the universal distance on that line by the placing of an object.

If the object is square and its sides are the same size as the unit of distance for measuring along the line, as in the top case in Figure 17, then d will always be 2 units of distance. Here we refer to the 2-unit deviation as a single d. Note that if an object with, say, shape 3×1 is placed on the line lengthwise, then, as in the middle case in Figure 17, d for negotiating that object will always be 2 units regardless of the length of the unit, because the trip between the two deviation units is parallel to the original line. If, however, the 3×1 object is placed orthogonal to the line of movement (see the bottom case in Figure 41), then a further two units of distance, that is one further d, will be added for every parallel line blocked by the object.

Figure 18 illustrates the principle of centrality: if we want to place a cell (the light cell top left) between two existing cells (dark), does it make a difference where we place it? The answer (mid-left) is that the more peripherally we place it, the less the increase in universal distance, and the more centrally we place it, the greater the increase. It follows (bottom left) that if we place cells evenly along lines, the increase in universal distance is greater than if we make some gaps large and others small. It also follows (mid-right) that an object placed in the centre of a space will increase universal distance more than one place towards the edge (because the effect on two dimensions will be the sum of linear effects). The principle of extension also follows: if we place a block on a longer line, it increases universal distance more than if we place it on a shorter line. Figure 19 illustrates the principle of contiguity: cells joined contiguously increase universal distance more than those placed discretely. Finally, Figure 20 illustrates the principle of linearity: contiguous cells arranged linearly increase universal distance more than if they are placed compactly. It should be noted that these principles interact. For example, in a finite space placing one cell contiguously with another will locally

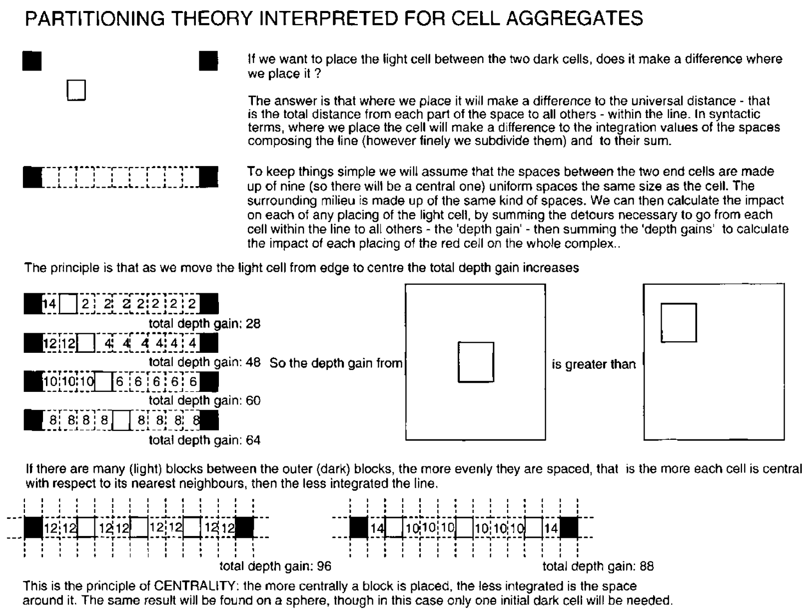


Figure 18.

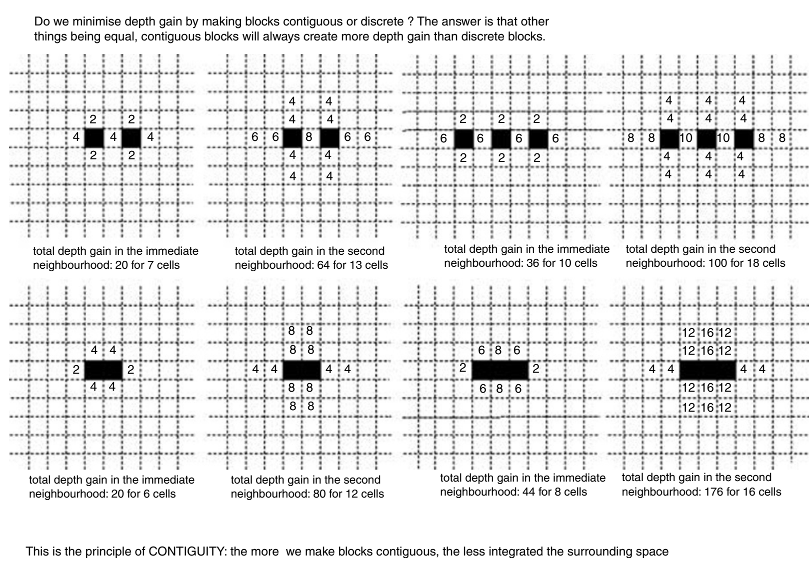


Figure 19.

increase distance but will also expand the space between the composite object and the boundary, thus reducing universal distance.

The impact of these laws on grids can be explored by constructing experimental grids made up of metrically uniform cells (they can be as small as we like, as long as they are uniform), and calculating the mean universal distance, or mean trip length, for each. Figure 21 sets out a series of experiments with grids each with 301 metrically uniform cells. The cells are circular in order to avoid the effect of corner joins. Each grid thus has the same number of metric cells and therefore the same number of distance elements. Differences between grids are therefore purely to do with the rearrangement of the cells into different configurations. In some cases, the rearrangement has left one cell that cannot be located in the grid. In each of these cases the cell has been added to

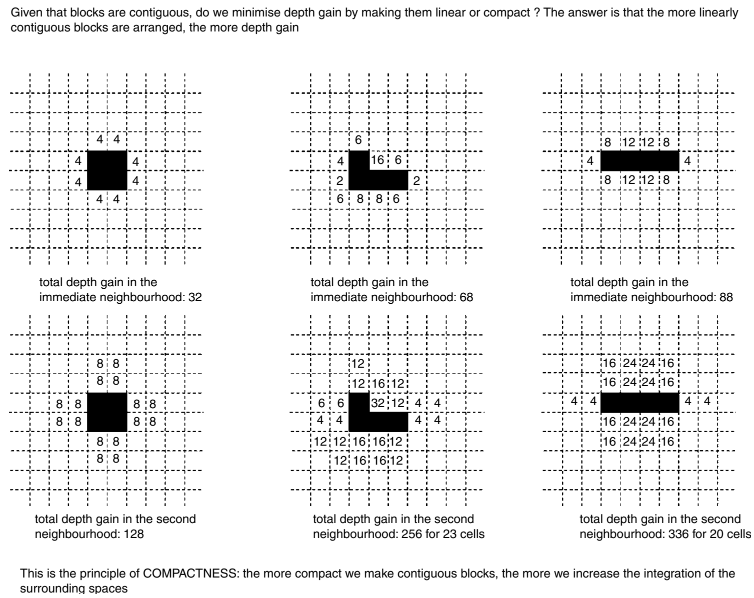


Figure 20.

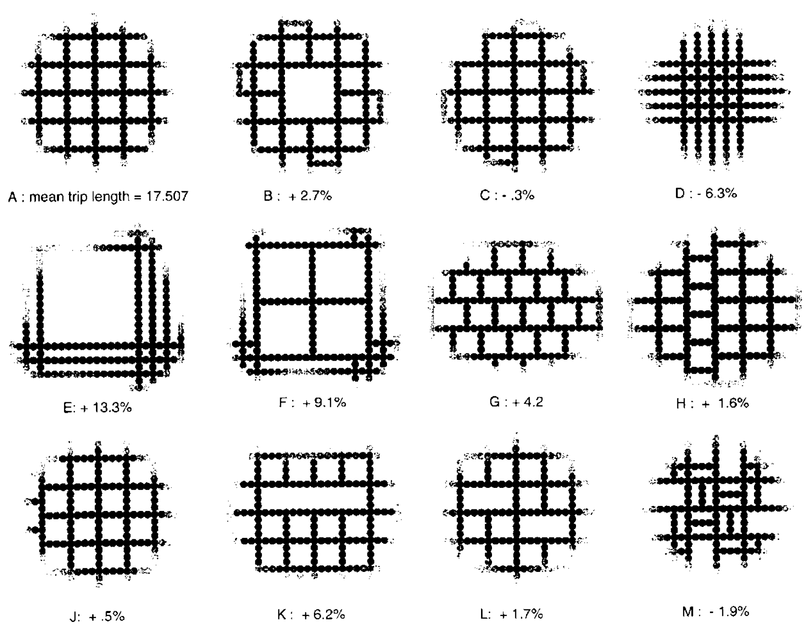


Figure 21.

the same position in the grid, namely the intersection of the third column (counting from the left) and the third row (counting from the top). Experiments with the sensitivity of the grid to the addition on one overlaid cell show that an additional cell overlaid in the centre of the uniform regular grid reduces the mean universal distance by 0.1% (it will of course slightly increase the total since there is an additional cell), while overlaying it on a corner cell increases it by 0.2%. These differences are then one or two orders of magnitude less than the effects of configurational changes below, and so can be discounted.

Using the regular uniform grid (Grid A) as the benchmark, we can then vary the configurations of grids to illustrate the effect of the four principles. In Grids B and C for example, we illustrate the 'centrality' principle by placing a block initially in the centre and then in the corner, while standardising the layout of the cells displaced from the centre. Placing the block in the centre increases the universal distance of the grid by 2.6%, while placing it in the corner reduces it by 0.3%. In Grid D, we take this further by reducing the scale of blocks in the centre at the cost of increasing them at the edge (a common form in the centre of towns, as noted in 'Centrality as a process' (Hillier, 2000)). The mean universal distance is reduced by 6.3%. If we do the opposite and make the centre block as large as possible, and place the small blocks at the edges (the number of small blocks remains the same, as in Grid E), we increase the mean universal distance by 13.9%, making a total difference between Grids D and E of just under 20%. In Grid F, we take Grid E and create a cross link through the centre. The effect is to increase mean universal distance by 9.1% compared to Grid A, but to reduce it by nearly 5% compared to Grid E.

We then illustrate the principle of extension. In Grid G we displace each vertical segment of cells between grid intersections one cell to the right and then to the left on alternate rows. We thus shorten all internal vertical lines with more or less neutral effects on block sizes. The effect is to increase mean universal distance by 4.2%. In Grid H, we break all horizontal lines close to the centre vertical, creating a pair of lines of fairly equal length at each level. The increase in mean universal distance is 1.6%. However, when we break the horizontal line near the edge vertical in Grid J, thus keeping some lines as long as possible at the expense of others becoming much shorter, the mean universal distance increases by only 0.5%, three times less than with a more central break in the lines.

The principle of compactness is illustrated in Grid K by converting the square central block of Grid B into a linear block of equal area. The effect is to increase the universal distance by 6.2% compared to 2.6% for the square block. We then illustrate the principle of contiguity by splitting the linear block into two in Grid L. The increase in universal distance is 1.7% compared with Grid A, but of course it is nearly four times less than for the contiguous linear block.

These grids are illustrative of course rather than a proper test, because huge combinatorics are involved, and in complex situations the four principles will interact. For example, in Grid M, we break many lines, and also make many smaller blocks in the centre. The result is a decrease in universal distance of 1.9% compared with Grid A in spite of the shortening of lines.

The law of centrality

It is now proposed that these four principles can be reduced to two formally demonstrable laws: a law of centrality and a law of compactness. The law of centrality proposes that an object placed centrally in a space will increase universal distance more than one placed peripherally. Consider again the line of n+1 cells with an object placed somewhere along it leaving n cells in some distribution on the two sides of the cell with at least one cell on each side. Wherever we place the object, D for one side of the line must be equal to D for the other, since each cell acquires one d for each cell on the other side of the object. (see Figure 18 mid-left). For example, if there are x cells on one side of the line and y on the other, then on one side D will be $x*y$ and on the other $y*x$. To establish D then, we need only establish it for one side of the line, since we may then multiply by 2 to get the total for the whole line. We therefore work by calculating D as the sum of d's for one side of the line.

Suppose then that the object is placed centrally on the line. It will then have equal numbers of cells on either side. Let m $(=(n-1)/2)$ be the number of cells on each side of the object. Each of m cells on one side then requires one deviation to go to each of cells on the other, giving a total of $m*m$ or m^2 deviations for each side. The total deviations, D, for the line with a centrally placed object, c, is then $2(m^2)$ or m^2 for each side:

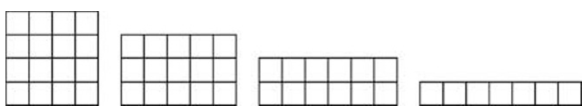
$$D(c) = 2m^2 \quad (1)$$

Now move the object one cell sideways. The total deviations for one side will then be $(m-1)(m+1)$ and for the other $(m+1)(m-1)$ or $2(m-1)(m+1)$ for the whole line. Now $m^2 > (m-1)(m+1)$ is a necessary inequality, as for example $3^2 > 2*4$ or $4^2 > 3*5$. Similarly, $(m-1)(m+1) > (m-2)(m+2)$ is a necessary inequality, as $2*4 > 1*5$ or $3*5 > 2*6$. In general: for $D(c_1c_2\ldots c_n$ representing steps away from the central location

$$m^2 > (m-x)(m+x) > (m-(x+1))(m+(x+1)) \quad (2)$$

with $x=0$ for the central object case $(m-0)(m+0)=m^2$. It follows that the greater the x, that is the farther the object from the centre, then the smaller is the product of $(m-x)(m+x)$. In other words, the farther the object from the centre, the lower the total D.

One way to think of this is geometrically. The perimeter of a rectilinear shape is the sum of its sides. Holding the perimeter of the shape equal, the area of the shape, that is the product of its longest and shortest side, is maximised when all sides are equal and reduces as we shorten one side and lengthen the other, eg 4*4 = 16, 5*3 = 15, 6*2 = 12 and 7*1 = 7 as in



In a sense, then, the law of centrality replicates the behaviour of area–perimeter ratios, even though we are dealing with linear effects.

In practical terms, this means that in a growing cellular aggregate such as a settlement, when faced with a choice of placing an object somewhere between two other objects, we should always place it close to one object and far from the other if we wish to minimise the gain in universal distance in the system as a whole. This means that gain-minimising decisions will always tend to create long and short lines rather than lines of similar length. This is clearly the case where a partitioning is made along an existing line, so that the two newly created sub-lines are colinear (ie share the same alignment). However, it is clear that it will also be the case for lines that are not colinear. The lower depth gain from a long and short line is not created by the rule but by the situation created by the rule. It has arisen from the intrinsic properties of a longer and shorter line compared to a pair of equal lines, and in fact the lower depth gain from the long and short line in fact results from the existence of the longer line, even when offset against a shorter line which was the by-product of its creation. If longer lines are beneficial even when offset by a colinear short line, it follows that a longer line will be beneficial anywhere even when it is not so offset by a colinear short line. It follows that to minimise depth gain in a system we should always conserve longer lines at the expense of shorter lines. This is the 'principle of extension' – always conserve long lines and partition shorter ones – and it thus follows as a corollary of the law of centrality. A second corollary is that placing two objects equidistant from each other and from other objects will increase universal distance more than placing them either close to each other or close to other objects, since the former will create many equal short lines, while the latter will create some longer and some shorter lines. In general, we may say that placing objects in proximity to each other increases universal distance less than placing them farther apart[12].

The law of centrality thus addresses the fundamental spatial problem of settlement: how to aggregate built forms in such a way as to preserve the interaccessibility which is potentially interrupted by those built forms, and how to maintain this as the settlement grows. It leads to a fundamental idea in the generation of settlement: that to minimise universal distance in the system (ie to maximise metric integration) the fundamental strategy must be always to conserve longer lines, if necessary at the expense of creating other short ones.[13]

The law of compactness

The law of compactness proposes that the more compact an object or group of objects, that is, the more its shape approximates a circle (or for

[12]This may be demonstrated with greater clarity in an unbounded system such as a torus. Consider two objects placed on the surface of a torus. The two objects have two distances from each other: a distance from one face of the object to the nearest face of the other object; and a second distance from the opposite face of the first object to the opposite face of the second object 'the other way round' the torus. These distances may either be equal or different. If they are equal – that is the objects are the farthest possible distance apart in any direction – then the law of centrality shows that the increase in universal distance is maximal. The more the objects are moved together to create a nearer and a farther distance, the more we have a shorter and a longer distance and therefore the less the increase of universal distance. It follows that placing objects close to each other in an unbounded system increases universal distance less than placing them farther apart.

[13]Also variation in the scale of the attached space will have substantial effects on the axial map.

practical purposes a square), then the less the increase in universal distance in the surrounding space. This may be shown by first considering the effect, as before, of placing an object on a line of n+1 cells. We know that the maximum increase in universal distance for each side is m^2 ($m=(n-1)/2$) for the case where the object is placed centrally. If we then place a discrete object on a second line with at least one line between the new and old line, then the gain on the second line will also be m^2, since the objects do not affect each other (see Chapter 8 of *Space is the Machine*, for a discussion of the case where lines are neighbours). In general, the depth gain for single discrete objects placed on distinct lines will be $2(m^2)$ or $n(m^2)$, where n is the number of lines. The rate of increase is therefore linear.

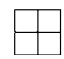
Now suppose that the objects are placed contiguously on neighbouring lines. This creates a more complex situation in terms of depth gain, which is illustrated in Figure 20. As we can see, depth gain is least at the edges and greatest in the centre. With m being the length of the line blocked and n the length of the partition (= the number of lines blocked), the depth gain can be calculated by the finite series

$$D = n^2m^2 + (n-2)^2m^2 + (n-4)^2m^2 \ldots (n-n)^2m^2 \qquad (3)$$

which gives a third-order polynomial function for the increase in universal distance with either increased partition length or line length. It can then be compared to the linear rate for discrete cells. If blocks are discrete, then universal distance increases linearly, and if contiguous the increase is a third-order polynomial function with increasing contiguity. This demonstrates the old ‘principle of contiguity’. However, as we will see below, we must also unify this with the idea of compactness.

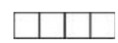
Consider the effect of an aggregate of objects forming an overall shape placed on a regular grid of lines. The shape will increase universal distance in two directions in the grid, which we can think of as horizontal and the vertical. Holding m, the length of the line on either side to the shape, constant, the increase in universal distance in one direction will be a third-order polynomial function of n, the number of contiguous cells composing that face of the shape. Alternatively, we can hold n constant and vary m, with the same result. These calculations will not be affected by the number of cells on the adjacent side of the shape, since these will only increase universal distance in the other, orthogonal direction. The overall increase in universal distance resulting from the imposition of the shape of the grid will then be the sum of the effects on each direction of the lengths of the two different faces of the composite object blocking that direction calculated by formula (3) applied independently to both directions.

Suppose then that the sides are equal, that is the object is maximally compact, say 2×2:

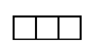


Holding m constant at, say, 3, the gain in universal distance will be $2(n^2m^2) = 2(2^23^2) = 72$ for each direction (made up of the two half-lines), or $4(n^2m^2) = 144$ for the whole object.

Now alter the shape of the object to a 1×4:



The gain in the vertical direction will now be $(4^23^2)+(2^23^2) = 180$ for each half-line, *2 for the pair of half-lines = 360. That in the horizontal direction will be $2(1^23^2) = 18$ for the pair of half-lines. The total gain is then 378 compared to 144 for the square object. In fact, if we reduce the object to a linear block of three cells:



then we have $2((3^23^2)+(1^23^2)) = 180$ for the vertical direction and $2(1^23^2) = 18$ for the horizontal direction, giving 208 which is still greater than 144.

The reason for the increase is simple. Since m is constant, n is the only variable in equation (3). When the block is square then $D = 2n^2$. However, if we replace the square object with a rectangular object, say, $(n-1)$ on one side and $(n+1)$ on the adjacent side, then all we need to know is the relation between $(n-1)^2+(n+1)^2$ for the two unequal half-lines of the rectangular object and $2n^2$ for the two equal half-lines of the square object. Since $(n-1)^2+(n+1)^2 = (n^2-2n+1)+(n^2+2n+1) = 2n^2+2$, it follows that $(n-1)^2+(n-1)^2 > 2n^2$ and that in general

$$(n-x)^2+(n+x)^2<(n-(x+y))^2 +(n+(x+y))^2 \quad (4)$$

From this it follows that a compact form will always generate less depth gain than an elongated form of equal area, and that the difference increases rapidly with increased elongation. As with the law of centrality, a simple geometrical idea underlies the law of compactness.

Impact of the laws on the basic generative process

How then do the spatial laws impact on the basic generative process? We have already seen that the social forces driving settlement formation are dual, with a residential component, driven by socio-cultural forces, and a public space component, driven by micro-economic forces. These correspond to a duality in the settlement form itself, with the invariant deformed wheel global structure formed by the public space process and the culturally specific interstitial local background areas formed by the residential process. We also note that there is a duality in the spatial laws, in that the compactness law addresses the physical component of the settlement, that is the size and shape of aggregate objects (ie blocks), while the centrality law addresses the spatial component, that is length of lines, distance of objects from each other, and so on. We recall that the output of the basic generative process in Figure 12 was deficient in both respects: blocks were overly varied in their shape and lines were insufficiently varied in their length. Our task was to explain the differences between the computer-generated model and the real cases by showing how the dual social processes impacted on the basic generative process through the intermediary of the spatial laws.

Two conjectures can now be proposed. The basic generative process guaranteed interaccessibility but it did not specify its degree or type, that is, it did not specify a more or less integrated process or a particular local geometry. To control this, one would need in the first instance to set a parameter for the compactness law regulating the size and shape of blocks, by specifying, for example, for how long and where one could continue adding to an existing block and when a new one had to be started. Such a parameter would in effect specify how the compactness law would influence the pattern and degree of universal distance in the background structure of the system in general. The first conjecture is that it is this local interaccessibility parameter controlling the generic block structure and operating through the compactness law that is set by the residential process and its socio-cultural drivers. It is through this that the characteristic local geometry of space is created in the first instance in the background residential areas of the settlement. Where this is set differently by different cultures, we find the kind of differences noted in the different parts of Nicosia (Figure 11). Where it is more homogeneous, we generate the kinds of generic regional differences in axial geometry that are indexed in the geometric and syntax values (Table 1) set out earlier.

The second conjecture is that with the growth of the settlement (and in fact in quite early stages) the public space process, led by micro-economic activity, sets a global interaccessibility parameter working through the centrality law. Since micro-economic activity is by nature integrative, this is not a variable, but a constant. Its effect is always to seek to conserve longer lines and to use these to minimise universal distance in the larger scale system. Since the effects it seeks are spatial, it operates directly on space and therefore works through the centrality law. The public space process thus tends to generate the local-to-global deformed wheel structure at whatever level of the settlement it is applied, including, where it is operative, local area structures. However, this is not all the micro-economic process does. In its loci of most concentrated activity it will generate not a linear system that minimises universal distance in the system as a whole, but a locally intensified grid that minimises movement from all origins to all destinations in the local region (see, for example, the central area of Konya in Figure 10) (Hillier, 1999a, b).

Looking at Konya, we can now see the settlement plan in a new light. We can see how spatial laws driven by the dual process have created the key features of the layouts: a deformed wheel global structure, an intensified grid forming the hub of the wheel, and the background of residential areas. However, there is an important respect in which the processes that create these patterns can be seen as a single process. The operation of the centrality law is *dual*, in that it creates both integration and segregation. In this it is capable of reflecting *in itself* the fundamental duality of the socio-cultural and micro-economic processes. The socio-cultural process, which creates the larger areas of background space in the city, is always a matter of imposing some restriction on integration and the natural co-presence that follows it through movement, while the micro-economic process operates of necessity by always maximising integration (minimising universal distance) in order to maximise natural co-presence in its spaces. The micro-economic process therefore naturally occupies that part of the duality of the law of centrality which generates the longer lines and the essential structure of the settlement, while the socio-cultural process equally naturally occupies the obverse side, the production of a larger number of shorter lines which construct the less integrated background of mainly residential space in the interstices of the global structure. Through the dual nature of the centrality law, then, the dual process acquires a single expression.

These conjectures require of course a whole research programme to test them, involving both simulations of settlement growth and on the analysis of real cases. However, some useful preliminary indications have been gained by some simple experiments with the impact of the centrality law on the basic generative process (using at this stage a manual process). For example, once we know the law of centrality, we can use it to maximise universal distance in a restricted random process by having a rule which requires the blocking of the longest line whenever an opportunity presents itself. Figure 22 is a manually generated outcome from applying this rule within the basic generative process. The outcome pattern is primarily composed of short lines, and (for the reasons given earlier) has very poor and unurban local to global synergy of .147, about as low as it can get for a small system). It also lacks the kind of global structure typically found in settlements, although it does begin to show signs of an interesting, but overly peaked, log distribution of line lengths (Figure 23). In short, it shows little sign of the spatial invariants of settlement we are looking for. In some ways, it is the opposite. Suppose then that we use the centrality law in the opposite direction, and set up a rule that forbids blocking a line once it has acquired a length of, say, five cells. This generates a pattern of many more long lines, as in Figure 24, which does have a good synergy score (0.820), but the lines do not construct a deformed wheel pattern with interstitial local areas, and as Figure

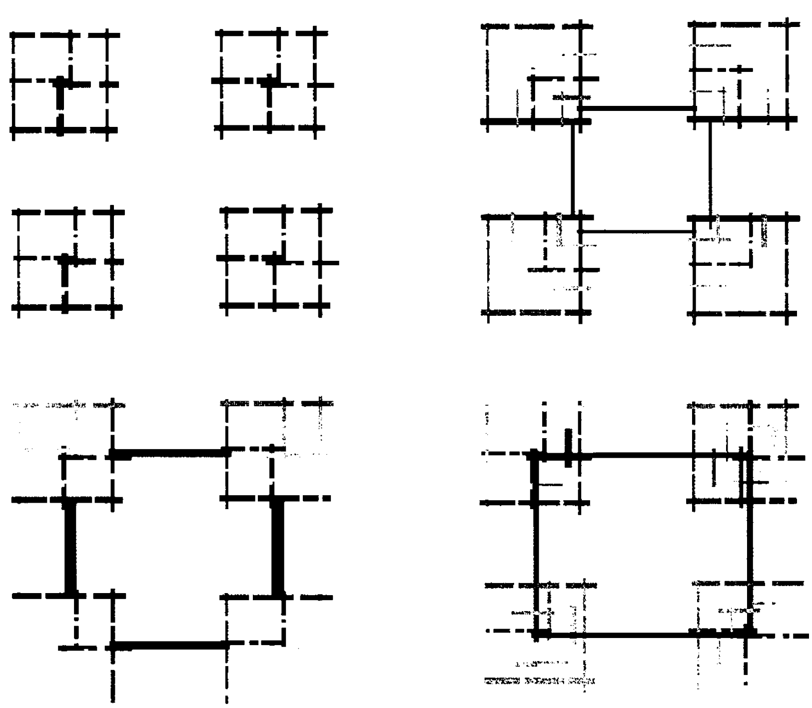


Figure 22.

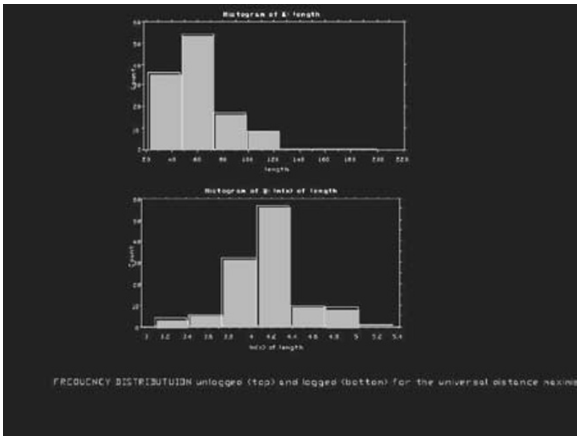


Figure 23.

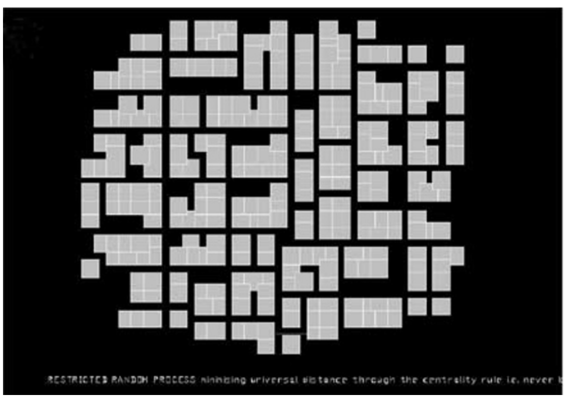


Figure 24.

25 shows the approximation of a log-normal distribution is quite poor.

Suppose then we apply the centrality law in the simplest and most localised way by setting up a rule that says that wherever you are adding a built form to the aggregate you have to choose a local location which preserves the locally longer line, but at the cost of continually creating shorter lines. Figure 26 is an outcome of such a process. Its global structure is overly biased towards the central horizontal line, but it is centre to edge, and the local areas are insufficiently structured in relation to the global core (giving it an urban synergy score of 0.729), but it does even at this stage of growth begin to look more like the log-normal distribution of line lengths, as in Figure 27.

This suggests that it may indeed be the duality of the centrality law in creating many shorter lines to compensate for each longer one that is in the last analysis responsible for the log-normal distribution of lines lengths in real settlements. However,

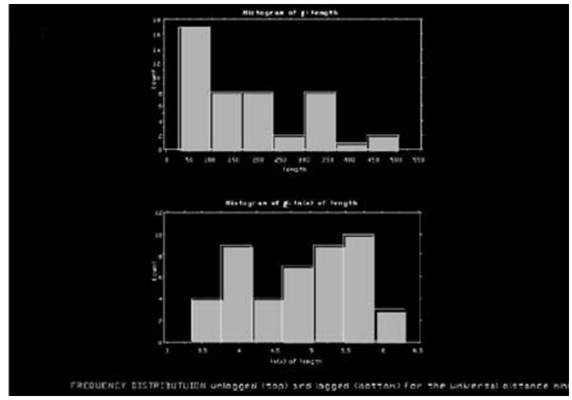


Figure 25.

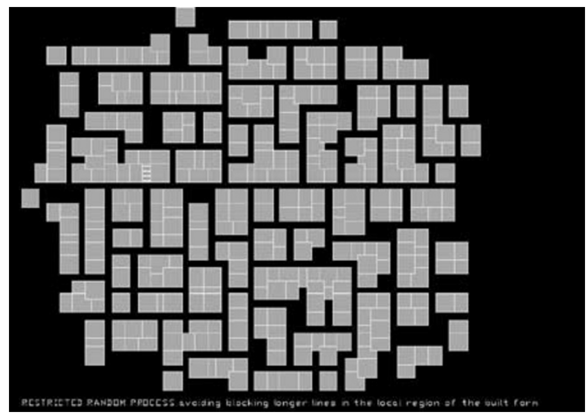


Figure 26.

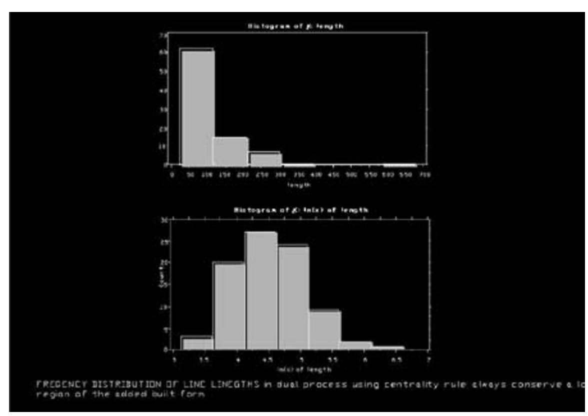


Figure 27.

although the tendency of the micro-economic process to use the longer line output of the centrality law seems to be invariant, the relations between these two aspects of the dual process should perhaps be seen as a variable. Sometimes, for example, the zones of background residential space seem to be no more than the by product of the micro-economic process, while elsewhere – Konya would be an example – there is a conscious parametrisation of the obverse side of the dual processes to create quite substantial regions of the urban grid, sometimes quite distant from the main settlement structure. In other cases, such as London, we find the local areas are much more closely related to the global structure, more axially integrated into it, and themselves have local to global deformed wheel structures.

This kind of variation suggests a rudimentary typology of settlement forms based on the different balance between the micro-economic and socio-cultural forces. Where the economic process is dominant from the beginning, we find linear or cross-road settlements and these are usually found on major routes between larger towns, a linear town being 'global structure only'. A deformed grid town is one in which both processes run in parallel. A regular orthogonal grid town is one in which the local cultural process is in the spatial image of the global economic process, as in mediaeval planted towns or early American towns, and where the whole grid is essentially a micro-economic rather than socio-cultural creation, as can reasonably be said both of mediaeval planted towns and early American grids[14].

We may then be within striking distance of grasping aspects of the pervasive logic by which apparently different social forces generate invariants in their settlement patterns as well as the more obvious differences. The key issues are the parametrisation of the cultural process which defines the local spatial geometry, and the balance between this and the emerging micro-economic process as the settlement grows. In the early stages of growth, the local socio-cultural process guarantees interaccessibility in the emerging settlement pattern but little more. It sets a parameter which by deciding the degree or ease of interaccessibility (ie more or less universal distance) specifies the local geometry of the settlement, covering both line length, angles of incidence and block size – all factors in interaccessibility. With growth, the universalistic and therefore globalising micro-economic process increasingly interposes on this process a simple depth minimising mechanism for each built form placing decision: conserve long lines, if necessary at the expense of creating many shorter lines. This will have the effect of generating a pattern of a few long lines and many short lines, and because the choices are regional this will be the case at every level – that is, this process will generate the pervasive log-normal distribution with a few long lines and a large number of short ones at every level. Changes in this fundamental pattern of growth will reflect essentially the changing balance between micro-economic and cultural forces, and this may (as historically in London) alter with the passage of time, with each alteration leaving its mark on the settlement geometry.

[14]Such 'economic' grids need however to be distinguished from the grids of administrative, garrison or ceremonial towns which characteristically are not pure grids but interrupted grids in which many lines, including some major lines, are interrupted by the facades of major public buildings at right angles.

However, the core issue is that the inherent duality of the spatial law of centrality is able to reflect the duality of these potentially conflicting social forces, and turn what is initiated as a dual process into a single process by which the locally highly differentiated and globally highly structured pattern of urban space come into being.

A reflection

The deformed wheel structure with its interstitial areas – the classic, although not the only, urban form – seems thus to be a product of an essentially metric process, optimising metric integration in some aspects, restraining it in others. Some may have noticed that this leads to a difficult question. Why should we continue to regard axial maps as topological structures, to be analysed through their graphs, when we have shown that they are generated through an essentially metric process? Would we not be likely to arrive at a better picture of the city if we subjected the axial map to metric analysis? It has already been suggested that the intensified grids found in centres and subcentres are best understood through metric integration analysis (Hillier, 2000). Is it not time to subject the axial map as a whole to such an analysis, or at least to a metrically sensitive analysis? In this way, we could surely counter one of the main objections to the axial map as a basis for graph analysis: that the nodes of the graph represent unequal elements.

The problem is that as soon as we introduce a metric dimension to an axial map – whether by using an analysis based on metrically uniform elements, or by weighting, say, line segments for length – configurational analysis produces not an enhanced version of the kind of picture given by the line graph analysis, but a very different picture: one that essentially picks out geometric centrality in the system, as we saw when we used more or less uniform line elements in the pseudo-system shown in Figure 15. If we applied this to a city like London, it would have the effect that a short alley off Oxford Street would seem to be more integrated than, say, the Holloway Road. In one sense it is of course, since it is closer to the geometric centre of London. However, in a more important sense we would seem to be losing one of the most important aspects of the integration analysis of the urban system: the substitution of a picture of geometric centrality by a picture of centrality in the line topology, one that identifies geometric centrality but then draws it out towards the edges of the system in all key directions, and even including parts of the edges of the system.

The question is then: which is the true picture. Is the one brought to light by the radius *n* analysis of the line graph in some sense identifying properties that are truly of the nature of the urban system and essential to its functioning? One thing is clear. Metric analysis of a large-scale system is very much poorer in its capacity to postdict the movement structure. In experiments carried out in 1986 (Hillier *et al*, 1986) on axial maps whose segments were weighted for length and used as the units of analysis, this very propensity to assign too high a movement prediction to lines adjacent to strong lines and too low a prediction to syntactically stronger but more remote lines, destroyed the normal approximate agreement between integration and movement. This suggests that the axial map, analysed as a line graph, might after all be capturing something that is of the essential nature of the urban system.

What can this be? There are two aspects to a possible answer: one substantive and to do with urban reality, the other cognitive and to do with how we interact with urban reality. Substantively, the empirical effect of the line inequalities in the urban system is to create a disjunction between geometric centrality in the system and topological centrality in the line map. In effect, centrality is topologically stretched from the geometrical centre to form links with the edge in all directions. In doing so it also structures the object by creating a relation between the local and the global organisation. The benefits of these are obvious enough: strangers are provided with easy-to-read routes from edge to centre and out again, and the system acquires local to global intelligibility and synergy. In contrast, it is easy to see that a system without the line inequalities in the right place and of the right type will degenerate into a labyrinth.

In fact, in terms of the micro-economic processes that create the deformed wheel structure, we find an even stronger argument when we consider the settlement not in isolation but as part of the wider system of settlements. Figure 22 illustrates what in Chapter 9 of *Space is the Machine* was called the 'paradox of centrality'. On the left are three notional settlements, each with its own internal integration core. But when (on the right) we join them into a single system and analyse the integration pattern for the system of settlements, we see that integration shifts to the edges of the settlements. Clearly, if we consider each settlement on its own, then the internal pattern of integration will approximate the internal movement structure, while if we consider them as a system of settlements the edge pattern will reflect movement in the overall system.

This is of course exactly what happens in real settlements. Movement patterns invariably have a local aspect and a global aspect, the former reflecting circulation within the system, the latter movement in and out of the system. Insofar as movement is driven by the micro-economic process, it generates both the intensified local grids of the centres and subcentres by reflecting the need to minimise distance from all points to all other points within the zone, and the linear links from the local to the global scale of the settlement, reflecting the need to minimise distance from certain points to certain others at the larger scale (Hillier, 1999a, b), including into and out of the system. Over time, this tension between the internal and external movement economies of the settlement is the fundamental reason why centrality tends to shift towards the edges of the settlement, unless strenuous efforts are made to inhibit it.

The deformed wheel structure is the key mechanism for this inhibition. In Figure 22, we draw four national settlements, each with a 'deformed wheel' case linking edge to centre (top left). Top right, we connect the edge lines of the four. The core goes to the edge of each. Bottom left, we link centre to edge lines to each other. The core penetrates into, but not across, the settlements. Bottom right, we extend the centre to edge lines into neighbouring settlements. The internal cores of each are nearly restored.

Cities as discrete geometries

The second reason why we might suspect that the axial map captures essential properties of the urban system is cognitive. The analysed axial map seems to approximate the intuitive picture we have of an urban system to an unexpected degree. A simple reason for this would be that human beings are excellent judges of simple linear distances when, for example, throwing a stone or a spear, or a ball of paper into a waste paper basket. However, this comparatively secure judgement of distance quickly breaks down when the system becomes nonlinear and involves changes of direction. This would make simple sense in evolutionary terms. Distance is a comparatively sophisticated and recent concept, and there is no obvious reason why we would expect to judge it as well in the highly nonlinear situations created by human settlement as when we are dealing with distance as a simple extension of bodily reach.

Complex spatial systems seem then to be dealt with cognitively through something more elementary. What might this be? The obvious candidate is discrete geometry: that we cognise complex spatial systems like cities as assemblages of interrelated geometrical elements rather than as complex patterns of metric distance (Goodman and O'Rourke, 1994). Discrete geometry is the application of the techniques of discrete mathematics such as graph theory to systems of discrete geometric elements, such as lines, convex spaces and visual fields. Space syntax, we now can say with hindsight (there was not much discrete geometry about when we started), is the application of discrete geometry to architectural and urban systems considering these first and foremost as systems of space.

If our cognitive representations of complex space are indeed discrete geometrical, then the strongest candidate as the element in the discrete geometry would be the line. Lines have the two key properties of being both very simple and very global. All we need to know is how far we can see from a point. Put more theoretically in terms of the city as a total visibility field we can follow Penn (and adapting Peponis's beautiful concept of informational stability as those regions in a spatial system that do not change topologically with movement (Peponis *et al*, 1997)) in arguing that a linear clique (a set of points which can all see each other) preserves informational stability for longest for moving individuals and thus offers the most economical – although not the most complete – picture of an overall system (Penn, 2001). Other discrete geometrical representations, such as visibility graphs (Turner *et al*, 2001) for example, give a much more complete account of the complexities of urban space, but it is not obvious that they would form the basis of a cognitive representation of the city as a whole. There is too much local information for the global picture to be clear. An axial map maximises local simplicity as a means to picturing global complexity. With visibility graphs, it is the other way round. Analysis of how we give directions in complex spatial systems (Hillier, 1999a, b) suggests that the axial maps may not be too far from the way we represent them to ourselves, that is as a matrix of lines where changes of linear direction are the key items of information that become organised into the whole picture.

If we intuit the spatial structure of the city as a *discrete geometry*, then it is reasonable that we should analyse it by treating the discrete elements as the nodes of a graph. We are tempted to add to this: that we represent the urban system to ourselves not simply as a discrete geometry, but as a simplified discrete geometry, in the sense that a series of near straight lines of the kind that are commonly found in cities (Hillier, 1999a, b) are internally represented as a line, so that the whole system comes to resemble an approximate grid. If this is the case, then it would be no more than a case of the 'imposition of a Euclidean framework on non-Euclidean inputs as argued by Petepan and her collegues commenting on (O'keefe and Nadel, 1978)

> …the hippocampus appears to impose a Euclidean framework on non-Euclidean inputs O'Keefe and Nadel (1978), who see in this process an instantiation of a Kantian *a priori* notion of absolute space…we propose that in 'distorting' the sensory inputs, theses spatial maps impose an order and a structure that

our spatial conceptual representations require …(Peterson *et al*, 1996)

Since it is also the line topologies that seem to correlate with movement in the different parts of the system, it seems hard to avoid the conclusion that the line representation of the city is not just a convenient simplification but something that touches the essential nature of the city. This does not mean that it cannot be improved or broken down more than it is now. However, it does seem likely that any future configurational analysis of the large-scale structure of cities will need to include some representation of its linear dimension as currently expressed, although perhaps crudely, in the axial map.

With or without the axial map, this account of how urban space is generated has unavoidable implications for how we model the city. Models in the past have used the fundamental concept of mass and the Newtonian mathematics of gravitational attraction as the guiding theoretical entities. The integration equations play the same role in configurational models as the Newton equations do in attraction-based modelling. But they do so on the basis of a discrete geometrical representation of the spatial structure itself, one that seems to engage the key role of visibility in how we cognise and interpret our surroundings. This has a further implication: as Chiron Mottram has argued (Mottram, 2001), configurational models are light-based rather than mass-based: they reflect the world we see rather than the world of distance and mass. The question is: how far do these cognitive realities intervene in the functioning of the urban system? On the evidence so far, it seems unlikely that we can arrive at a theoretical model of the city without them.

Axial maps are to be credited as follows:

- Atlanta: Mark David Major
- The Hague: Laurie Neale
- Manchester: Polly Fong
- Old Hamedan: Kayvan Karimi
- London: Chang Hua Yoo
- Chicago: Mark David Major
- Shiraz: Kayvan Karimi
- Venice: Erica Calogero
- Teotihuacan: Ruben Garnica
- Apt: Ricky Burdett
- Serowe: Glenn Mills
- Paranoa: Frederico de Holanda
- Amsterdam: Xu Jianming
- Santiago: Margarita Greene
- Athens: Valentina Karvounzi
- Nottingham: Mark David Major
- Bristol: Kayvan Karimi
- York: Kayvan Karimi
- City of London: Julienne Hanson
- Baltimore: Shazir Shah
- Tokyo: Shinichi Iida & Masaki Nishibori
- Nicosia: Konstantinos Kypris

Acknowledgements

Thanks are due to Sema Kubat for permission to use the black on white map of Konya.

References

Bloom, P., Petersen, M., Nadel, L. and Garrett, M. (1996) *Language and Space*. Cambridge, MA: MIT Press.

de Holanda, F. (1977) PhD Thesis, Bartlett School of Graduate Studies, University of London.

Goodman, J. and O'Rourke, J. (1997) *Discrete and Computational Geometry*, CRC Press, New York.

Hillier, B. (1990) The architecture of the urban object, *Ekistics*, Special Issue on space syntax research, pp. 5–22.

Hillier, B. *et al*. (1993) Natural movement: or, configuration and attraction in urban pedestrian movement, *Environment and Planning B: Planning & Design*, **20**: 29–66.

Hillier, B. (1996a) *Space is the Machine*. Cambridge: Cambridge University Press, pp. 470, Paperback, 1998.

Hillier, B. (1996b) Cities as movement economies, *Urban Design International*, **1**: 49–60. Awarded AESOP prize for the best paper in a European planning journal of 1996.

Hillier, B. (1999a) The hidden geometry of deformed grids: or, why space syntax works, when it looks as though it shouldn't, *Environment and Planning B: Planning & Design*, **26**: 169–191. Theme Issue on Space Syntax Symposium. Also in Vol 3 of the Proceedings of the First International Space Syntax Symposium, 1997.

Hillier, B. (1999b) Space as a paradigm – Proceedings of the Second Space Syntax Symposium, Brasilia, 1999.

Hillier, B. (2000) Centrality as a process: accounting for attraction inequalities in deformed grids, *Urban Design International*, **3/4**: 107–127.

Hillier, B. and Hanson, J. (1984) *The Social Logic of Space*. Cambridge: Cambridge University Press, 281pp. Paper-back, 1989.

Hillier, *et al*. (1986) Spatial configuration and use density at the urban level: towards a predictive model. Appendix: The distance factor Unit for Architectural Studies, University College London.

Karimi, K. (1998) Continuity and change in old cities. PhD Thesis, Bartlett School of Graduate Studies, UCL, University of London.

Loumi, A. (1988) Spatial configuration of towns in North Africa. PhD Thesis, Bartlett School of Graduate Studies, UCL, University of London.

Martin, L. and March, L. (1972) *Urban Space and Structures*. Cambridge: Cambridge University Press, pp. 272.

Mottram, C. (2001) Personal communication.

O'Keefe, J. and Nadel, L. (1978) *The Hippocampus as a Cognitive Map*. Oxford: Oxford Univeristy Press.

Penn, A. (2001) Space syntax and spatial cognition: or why the axial line? Proceedings of the Third Space Syntax Symposium, Georgia Institute of Technology, Atlanta, 2001.

Peponis, J. *et al*. (1997) On the description of shape and spatial configuration inside buildings: convex partitions and their local properties, *Environment and Planning B: Planning and Design* **24**: 761–781.

Peterson, M., Nadel, L., Bloom, P. & Garreth, M: 'Space and language' in Peterson M, Nadel L, Bloom P & Garreth M *Language and space*, MIT 1996, 569.

Turner, A., Doxa, M., O'Sullivan, D. and Penn, A. (2001) From isovists to visibility graphs: a methodology for the analysis of architectural space, *Environment and Planning B: Planning & Design* **28**: 103–121.

Vischer-Skaburskis, J. (1974) Territoriality and its relevance to neighbourhood design: a review, *Journal of Architectural Research* **3** (1): 39–44.