THE GENETIC CODE FOR CITIES – is it simpler than we think ?

Keynote paper for conference COMPLEXITY THEORIES OF CITIES HAVE COME OF AGE at TU Delft September 2009

Bill Hillier

Bartlett School of Graduate Studies
University College London
Gower Street
London WC1E 6BT
U.K.
e-mail: b.hillier@ucl.ac.uk
www.spacesyntax.org (academic)
www.spacesyntax.com (consultancy)

ABSTRACT

Most recent mathematical characterisations of the city are statistical descriptions describing the distribution of spatial or functional properties of cities, but in abstract statistical space rather than real space. By definition this omits how elements with these properties connect to each other in real space, and so omits any account of the structure of the system. Here I show three things. First I show how statistical and other numerical characterisations of cities can be turned into structural characterisations. Second, I show that with this capability we can find a universal characterisation of certain deep or universal structures common to the spatial form of all cities. Third, I outline the 'genetic' process that gives rise to these universal structures in two phases: a spatial process through which simple spatial laws govern the emergence of characteristically urban patterns of space from the aggregations of buildings; and a functional process through which equally simple spatio-functional laws govern the way in which aggregates of buildings becomes living cities. This dual process is suggested to be akin to a 'genetic' code for cities.

Is there a universal city ?

On the face of it, cities seem to be made up of two very different things: a slowly changing, physical system, made up of of buildings linked by streets, roads and infrastructure; and a more rapidly changing human system made up of movement, interaction and activity. The human system seems to be superimposed on the physical system, and in some way to constitute its functional system, but has an unknown relation to it. The question every theory must address is: what, if any, is the relationship between the two systems ? How, we might say, does the city work as a socio-technical system ? Theoretical questions about cities then both look like classical structure-function questions, and questions about the relation between human beings and the physical world, so two difficult kinds of questions rolled into one.

On the face of it, the different rates of change of the physical and human systems would seem *in principle* to preclude all but the loosest of relations between the two. Again and again we find new patterns of activity fitting into already existing networks of space, with changes more in scale than structure - for example, the City of London has maintained a similar spatial network through an astonishing series of changes in its social and economic patterns over centuries. If the two systems don't co-vary, so that when one changes the other does, then surely there cannot be any kind of exact, and so quantifiable, relationship between the two.

But this does not quite exhaust the possibilities of finding exact and quantifiable relations. It is possible also that cities may have structural properties and structure-function interdependencies, which are both quantifiable and also universal in the sense that they are relatively indifferent to changes in social and economic circumstances, while reflecting both in a generic way. A theory of such a kind would be something like a theory of a *universal city* underlying cities in general. The aim of this paper is to present such a theory, using space syntax as a formal basis for the analysis

of spatial networks in cities, and suggesting that a theory of the universal city – what all cities have in common – is a necessary precursor the theories of specific cities or cultural or economic types of cities.

Space syntax is a formal way of looking at cities that sets out from the study of the network of space – streets and roads – that holds the system together, rather than from an assemblage of 'discrete zones', as is the usual practice (Wilson 2000). In what follows, we first use space syntax methods to bring to light some surprising regularities in the way in which all city networks are constructed, covering both the geometry and configuration of spatial networks, and functional as well as spatial phenomena. These regularities seems to be underlying 'structures' in that they lie below surface appearances, and are only brought to light in the laboratory, so constituting what Hacking (Hacking 1983) has called *created phenomena*. On the basis of these structures we propose a new universal definition of a city as a *network of linked centres at all scales set into a background network of residential space*. We then show that this universal pattern comes about in two interlinked but conceptually separable phases: a *spatial process* through which simple spatial laws govern the emergence of characteristically urban patterns of space from the aggregations of buildings; and a *functional* process through which equally simple spatio-functional laws govern the way in which aggregates of buildings becomes living cities. It is this dual process that is suggested can lead us in the direction of a 'genetic' code for cities.

Describing the network

But first we have the problem of describing the network of space. In the first instance, we will use the *least line maps* of cities developed by space syntax (Hillier & Hanson 1984) which are probably the simplest consistent representations of urban grids. These can in small scale cases be created algorithmically by using the UCL DepthMap software (Turner 2002, Turner, Penn & Hillier 2005, Turner et al 2006) but for large scale urban systems this is computationally prohibitive, so least line maps are commonly digitised using the rules for creating and checking maps set out in (Hillier & Penn 2004).

Analysis of least line maps for real cities brings to light some remarkable consistencies, common to both organic and more geometric cities. First, we find that at all scales, from the local area to the whole city, least line maps are made up of a very small number of long lines and a very large number of short lines (Hillier 2002), so much so that in terms of the line length distributions in their least line maps cities have been argued to have scale-free properties (Carvalho & Penn 2004). This is just as true of more geometric cities such as Chicago and Athens, as it is for more 'organic' (meaning lacking obvious geometry) such as Tokyo or London.

Line length distributions are of course a statistical property of cities, and in themselves say nothing about structure. But looking at the patterns formed by lines of unequal length in real cities, we find some even more remarkable consistencies, now of a geometric as well as metric kind. Looking at least line map of the - arbitrarily selected - section of Tokyo shown in *Figure 1*, the first thing the eye intuitively picks out are *line continuities*, that is lines joined by nearly straight connections. If we move along one of these we are very likely to find another at the end of the line, and then another. This tends to happens at more than one scale, and at each scale the lines are locally longer than lines which lack this kind of angular connection. Probabilistically, we can say the longer the line, the more likely it is to end in a nearly straight connection to another line.

We also see a much larger number of shorter lines with near right angle connections, forming more local grid like patterns. Again if you find one then there are likely to be several others in the immediate neighbourhood. We can also say the shorter the line, the more likely it is to end in a right angle or near right angle. These are the opposite properties to those we find in highly formal cities, like Brasilia or pre-Columbian Teotihuacan, where the longest lines end at right angles on the most important buildings.

Figures 1 and 2 Arbitrary sections of the least line maps of a section of Tokyo (left) and London (right)

In spite of the historic and functional differences between the two, we can make exactly the same two points about the section of London shown in *Figure 2*. In fact, we find it to be true of cities in general, geometric (where line continuities tend to be actually rather than nearly straight) as well as organic. Through the *geometry* and *scaling* of their street networks, cities acquire a kind of *dual* structure, made up of a dominant *foreground network*, marked by linear continuity (and so in effect *route continuity*) and a background network, whose more localised character is formed through shorter lines and less linear continuity. Looking across cases, this seems to be the generic form of the city.

Movement potentials as structures

What then do these patterns in the spatial networks of cities means ? and do they relate to function in any way ? We can take the next step by looking more closely at these structures using the DepthMap software, developed by the space syntax research group at UCL to analyse urban spatial networks in terms what we call *movement potentials.*

The basic element in DepthMap is the street segment between intersections. DepthMap generates this automatically from the least line map, and Space Syntax Limited has now developed algorithms to derive it from road centre line data (allowing whole regions, or even whole countries to be modelled). DepthMap allows 3 definitions of the distance between each segment and each of its neighbours: *metric*, that is the distance in metres between the centre of a segment and the centre of a neighbouring segment; *topological*, assigning a value of 1 if there is a change of direction between a segment and a neighbouring segment, and 0 if not; and *geometric* - assigning the degree of the angular change of direction between a segment and a neighbour, so straight connected are 0-valued and a line is a sequence of 0-valued connections, so that the linear structure of cities is captured. It then uses these 3 concepts of distance to calculate two kind of measure: syntactic *integration*, (mathematical *closeness* with the normalisations set out in Hillier & Hanson 1984)*,* which measures how close each segment is to all others under each definition of distance; and syntactic *choice* or mathematical *betweenness* (FOOTNOTE on how this has been measured in space syntax), which calculates how many distance-minimising paths between every pair of segments each segment lies on under different definitions of distance. So using the *metric* definition of distance we find the system of *shortest path* maps for integration and choice, with the *topological* definition we find the system of *fewest turns* maps, and with the *geometrical* definition we find the system of *least angle change* maps. Each of the 6 measures (2 measures with 3 definitions of distance) can then be applied with the 3 definitions of distance used as definitions of the *radius* at which the measures can be applied, giving a total of 18 measures, which can of course be applied at any radius, so yielding a potentially very large set of possible measures - for example least angle change choice at a metric radius of 800 metres – which would be infinite if we count the smallest variation in metric radius.

We think of *integration* as measuring the *to-movement potential* of a segment as a *destination*, since the measure describes its *accessibility* or *how easy it is to get* to from all other segments; and of *choice* as measuring the *through-movement potential* since the measure describes *how likely you are to pass through*

the segment on trips, and so its potential as a *route*, from all segments to all others. Since the selection of a destination and the selection of a route are the two prime components of any trip, we have then a well-grounded set of techniques for identifying movement related structural patterns in cities, and looking for functional correlates. In fact, using this panoply of measures, it was quickly possible to show that human movement follows *least angle change* paths and not shortest paths (the most likely explanation being that people use an angular geometric model of their environment to calculate distances), so the least angle change definition of distance is the default setting in DepthMap (Hillier & Iida 2005). Large numbers of studies have failed to suggest any reason why this should be changed. Similarly, metric radii have been shown again and again to be the most effective radius settings, with analyses typically being run across a range of metric radii, typically 250, 500, 750, 1000, 1250, 1500, 2000, 2500, 3000, 4000, 5000, 7500, 10000, 15000, 20000 metres and so on. So the standard measures used in DepthMap based studies are least angle integration and choice measures at variable metric radii. DepthMap also allows the weighting of segment by their length. So, for example, with the choice measure, a long segment with many buildings would generate and receive more movement than a short segment with few buildings, so the measure for the segment can be weighted by the products of the lengths of each origin-destination pair used in the calculation. In general, least angle measures with variable metric radii bring to light linear structures in the city, while metric measures with variable metric radii show a kind of area patchwork by finding discontinuities between areas. (Hillier, Turner, Yang & Park 2007). We will now look at some of the linear patterns identified by these measures.

Global structures

Applying these measures to real city networks, we again bring to light some remarkable consistencies by the simply procedure of colour banding mathematical values from red (dark) through orange and yellow to green and blue (light), meaning to strong to weak. For example, in case after case, least angle integration (normalised closeness, or to-movement potential) analysis without radius restriction (so the most 'global' form of the analysis), identifies a dominant structure in the form of what we call a *deformed wheel*, meaning a 'hub' of lines in the syntactic centres, strong 'spoke's linking centre to edge and strong 'rim' lines(closely reflecting the patterns brought to light by the earlier syntactic analysis of topological closeness of the least line map). *Figure 3* and *Figure 4*, for example, show the underlying deformed wheel pattern in both metropolitan Tokyo (with multiple rims) and London within the M25. Equally, the least angle choice (betweenness, or through movement potential) measure commonly identifies a network spread through the system, though strongest in the more syntactically central locations (see *Figure 5*).


Figures 3 and 4 showing least angle integration (normalised closeness) for metropolitan Tokyo (left) and London within the M25 (right) in each case showing a variant of the 'deformed wheel' structure, with multiple rims in the case of Tokyo

By combining the two measures, we can make powerful comparisons between the global structures of different cities, showing them to be variations on a common theme. For example, if we look at the combined global to and through movement potentials of London within the M25 we see a pattern in which there is a strong cluster of grid like spaces in the centre, and strong radials linking the centre to the edge, but very little in the way of lateral connections between the radials before the M25 (the 188 km ring road around London at a radius of about 30km). This

very much reflects what London is like. It is very difficult to go anywhere without going through the centre. If we look at the same analysis of Beijing we find almost the contrary: a relatively weak centre (in that we find strong intersections but not a grid like central area), and strong laterals (the ring roads) but relatively weak radials. By the way, the new business distinct is being developed exactly where these potentials are strongest on the east side. If we then look at Tokyo we find strong radial structure and strong lateral structure, and also a fairly strong and extended central area. This of course corresponds to Tokyo's pattern of sub-cities which occur were the radials and laterals intersect. Again in contrast, Suzhou we find the radial pattern but almost wholly confined to the historic area, reflecting the growth of that particular city as five relatively separates 'islands'. *Images to be shown at conference.*

These global patterns seem not to be confined to space alone, but seem to engage land use patterns, most notably the formation of local centres and sub-centres. For example, by setting the analysis of global least angle through-movement potentials in London within the M25 alongside Mike Batty's remarkable map of the London's 168 largest centres, we find a strong 'eyeball' correspondence. However, the image also makes clear that the global properties shown in the map are not sufficient in themselves to identify the location of centres. We typically find for example that along the length of a high global movement potential alignment we find the centre occurring only in certain locations. For example, if we take the Edgeware Road between the North Circular Road and Oxford street, there are three high streets with the rest fairly free of shops. In each case, the centre occur where local *grid intensification* (a dense and smaller scale local grid) co-incides with the globally strong alignment. *Image to be shown at conference.*
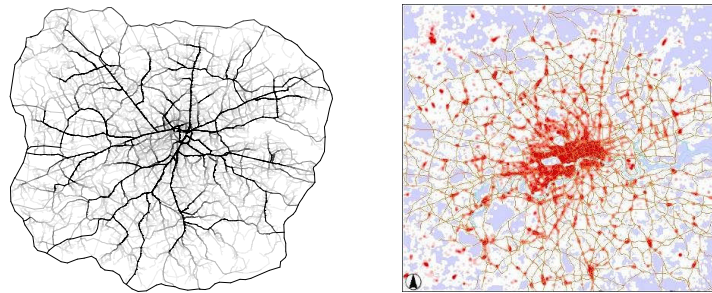
Figures 5 Least angle through movement potential with no radius restriction for London within the M25 (left) and Figure 6 Batty's map of the 168 main centres and sub-centres in the same area

Local structure

But local structures are also highly significant in their own right. For example, setting the least angle through movement measure to a radius of 750 metres, the analysis seems to identify in red

Figure 7 Least angle through movement potential at a radius of 750 metres in an area of north west London with the dark lines approximating the urban villages

(dark) most of the 'urban villlages' in a region of north west London, and, in general, by varying the radius, we usually find that there as some radius at which a local village or high street is identified as having the highest though-movement potential within the area defined by that radius: Marylebone High Street at 1250 metres for example, or the much smaller Lamb's Conduit Street (one of London's surprising village high streets) at 250 metres, or Bow Lane in the City of London at 200 metres. *Images to be shown at conference.*

But if we look more closely at London's – mainly linear – centres we find an even more striking regularity. In all of ten cases recently investigated, a particular segment or segments, usually at one or both ends of the high street, had the peak choice value at low radius (400 metres) then as the radius is raised contiguous segments along the high street were added, until the whole high street was covered, usually at a radius of about 2000 metres. This suggests that centre grow and work in a *multi-scale* way, in that different parts of the high street have different scale of reference to their urban context. *Image to be shown at conference.*

These effects are not confined to 'organic' cities. A recent study of the historic grid-iron pattern of Suzhou in China showed that the pattern of differently scaled centres could be identified by varying the radius of the least angle through movement measure. *(Images comparing the land use pattern to the spatial analysis will be show during the conference presentation).* Again in a recent study of unplanned areas in Jeddah in Saudi Arabia, least angle through movement potentials at a radius of 2.5 kilometres was able to pick out all the local centres, and by varying the radius from 1000 to 3000, it was possible to distinguish between the smaller scale centre in the more isolated northern parts of the area (1000 metres) from the larger scale centres closer to the Mecca Road in the southern parts. *Again, images will be shown at the conference presentation.* This kind of emergent pattern can even be found in Brasilia ! *Again, images will be shown at the conference presentation.*

It is vital that these apparent links between spatial and functional structure are only found with the least angle version of the analysis of through movement potentials. If we substitute metric distance for least angle distance and make the same analyses, the results are usually functional nonsense. For example, in the Jeddah unplanned areas, the metric version of the analysis finds none of the functional structures at low radius, and as the radius is raised it picks out a highly complex route, with dozens of turns, running from north-west to south-east and with no links to the pattern of centres. The reasons for this are not difficult to find. If we imagine a grid like structure and imagine a route diagonalising the grid, then a small change in the angle of just one segment would make that route either longer or shorter than the route around the edge of the grid. *Images will be shown in the conference presentation.* This is what we typically find: that metric analysis of through movement potentials, especially in more regular grids, finds seemingly arbitrary complex diagonals which have no relation to functional structure. For example, in Beijing, the metric version of the analysis (unlike the least angle version) fails to identify Changan Avenue (the 8 lane boulevard crossing Beijing from east the west and passing between the Forbidden City and Tianamin Square) as a strong alignment, and in London, while least angle analysis identified the major shopping street, Oxford Street as strongest, metric analysis finds Camberwell Green, a mid sized centre in south London. *Images of all these will be shown in the conference presentation*

A new definition of the city

The consistency with which these metric, geometric, configurational and functional regularities are found in superficially different kinds of cities in different parts of the world leads us inexorably to a new 'network' definition of the spatial form and functional form of the city. By some as yet unknown (but see below) process, cities of all kinds, and however they begin seem to evolve into *a foreground network of linked centres at all scales, from a couple of shops and a café through to whole sub-cities, set into a background network of largely residential space*. The *foreground* network is made up of a relatively small number of longer lines, connected at their ends by open angles, and forming a super-ordinate structure within which we find the *background* network, made up of much larger

numbers of shorter lines, which tend to intersect each other and be connected at their ends by near right angles, and form local grid like clusters.

This definition of the city entails a re-definition of centrality in cities. We call it *pervasive centrality* in that centrality functions diffuse throughout the network, at all scales, from the city as a whole to the local network of streets. The pattern is far more complex than envisaged in theories of *polycentrality*. It is notable also that pervasive centrality seems spatially sustainable because it means that wherever you are you are close to a small centre and not far from a much larger one. (Hillier 2009)

How then are these seemingly tight and generic relations between spatial and functional structures to be explained ? The answer may lie in two key new phenomena which research using space syntax has brought to light. The first we call *spatial emergence*: the network of space that links the buildings together into a single system acquires emergent *structure* from the ways in which objects are placed and shaped within it. This process is law-governed, and without an understanding of it the spatial *form* of cities cannot really be deciphered. How the city is physically *built* is critical. Cities are not simply reflections of socio-economic processes, but of the *act of building* in the light of these processes. The 'fact of the act' imposes a new framework of lawful constraints on the relation between socio-economic activity and space. The second is *spatial agency*: the emergent spatial structure *in itself* has lawful effects on the functional patterns of the city by, in the first instance, shaping movement flows, and, through this, emergent land use patterns, since these in their nature either seek or avoid movement flows. Through its influence on movement, the urban grid turns a collection of building into a living city. Movement is literally the lifeblood of the city.

It is these two linked processes of *spatial emergence* and *spatial agency* that set in train the self-organising processes through which cities acquire their more or less *universal* spatial form. These two processes are rendered more or less invisible by the standard method of modelling cities as discrete zones linked by Newtonian attraction. In the syntax approach to network modelling, the differences in attraction found in different parts of the network are *outcomes* of the self-organising process, and so theoretically (as opposed to practically) speaking, should not be taken as a given. But perhaps more than any other factor, it has been the - equally Newtonian ! - assumption that space can only be a neutral background to physical processes, rather than an active participant in them, that has rendered these space-based dynamics invisible to urban modelling, and so obscured the path from model to theory.

We will now look at *spatial emergence* and *spatial agency* in turn. In what follows it will be made clear that this is not a fully-fledged answer to the question of how we move from modelling cities to a genetic theory of cities. But we believe it is the first steps on the way. What is being shown, in effect, are new urban phenomena in need of a clearer and more unified mathematical formulation. The *theory* is pretty unified. Space syntax might be described as a mathematical patchwork set within a unified theory. What it needs is a unified mathematical treatment.

Spatial emergence

To understand the emergence of the spatial form of urban network we need first to understand its topology then its geometry. The basic form of all cities is one of discrete groups of contiguous buildings, or 'blocks', usually outward facing, defining a network of linear spaces linking the buildings. How can this arise? In fact very simply. If we take cell dyads (*Figure. 8*, top left), representing buildings linked by entrances to a bit of open space, and aggregate them randomly apart from a rule that each dyad joins its bit of open space cell to one already in the system (forbidding vertex joins for the buildings, since no one joins buildings corner to corner), a pattern of buildings and spaces emerges with the topology of a city - outward facing blocks defining a linking network of linear space - but nothing like its geometry, in spite of being constructed on a regular grid (Hillier & Hanson 1984). The 'blocks', and so the spaces, are the wrong shape. Where then does the characteristic urban geometry come from?
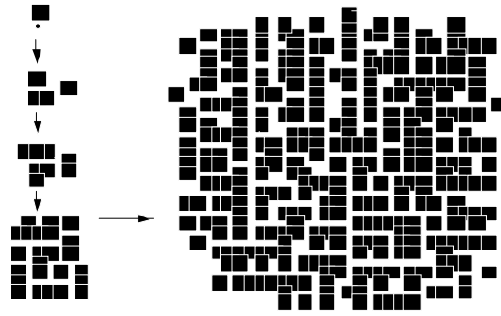
Figure 8 Aggregating dyads of open and closed cells by a restricted random process

To understand this we need first to think a little about the network of space in cities and how we interact with it, and the role that different notions of distance might play. Space in cities is about seeing and moving. We interact with space in cities both through our bodies and our minds. Our bodies interact with the space network through moving about in it, and bodily the city exists for us as a system of *metric distances*. Our minds interact with the city through seeing. By seeing the city we learn to understand it. This is not just a matter of seeing buildings. We also see space, and the city comes to exist for us also as a visually more or less complex object, with more or less visual steps required to see all parts from all others, and so as a system of *visual distances*. This warns us that distance in cities might mean more than one thing.

But we also need to reflect on the fact that cities are also collective artefacts which bring together and relate very large collections of people. Their critical spatial properties of cities are not then just about the relation of one part to another, but of *all parts to all others*. We need a concept of distance which reflects this. We propose that if *specific distance* means the common notion of distance as the distance, visual or metric, from *a* to *b*, that is from an origin to a destination, *universal distance* means the distance from each origin to all possible destinations in the system, and so from all origins to all destinations (Hillier 1996).Why does this matter? Because universal distance behaves quite differently from the normal metric and geometric concepts of distance that we use habitually. For example, if, as in *Figure 9* we have to place a cell to block direct movement between two cells, the closer we place it to one of the outer cells the less the total distance from each cell to all others will be, because more cell-to-cell trips are direct and do not require deviations around the blocking object.
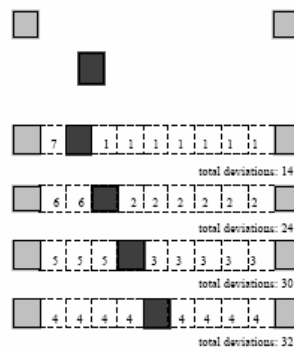

Figure 9 Moving an object between two others from edge to centre increases the sum of distances from all cells to all others

The same applies to intervisibility from all points to all others (*Figure 10*). As we move a partition in a line of cells from centre to edge, the total inter-visibilty from each cell to all others increases, though of course the total area remains constant. Both metric and visual effects arise from the simple fact that to measure inter-visibility or inter-accessibility we need to square the numbers of points on either side of the blockage.
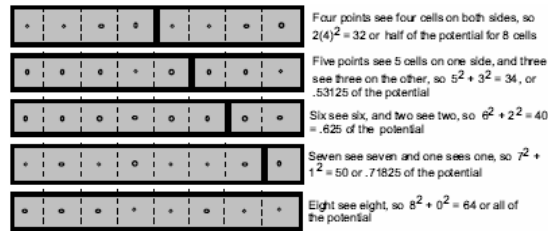
Four points see four cells on both sides, so $2(4)^2 = 32$ or half of the potential for 8 cells

Five points see 5 cells on one side, and three see three on the other, so $5^2 + 3^2 = 34$, or .53125 of the potential

Six see six, and two see two, so $6^2 + 2^2 = 40$ = .625 of the potential

Seven see seven and one sees one, so $7^2 + 1^2 = 50$ or .71825 of the potential

Eight see eight, so $8^2 + 0^2 = 64$ or all of the potential

Figure 10 Moving a partition from centre to edge increases total inter-visibility

So all we need to know is that twice the square of a number, *n*, will be a smaller number than $(n - 1)^2 + (n + 1)^2$ and that in general:

$$2n^2 < (n - x)^2 + (n + x)^2 \qquad (1)$$

We can call this the 'squaring law' for space. It applies when, instead of being interested in, say, the distance from *a* to *b*, we are interested in the distance, metric or visual, from each point in the system to all others. In space syntax these 'all to all' properties are called *configurational* to distinguish them from simple relational or geometric properties

So why does this matter? Because how we place and shape physical objects, such as urban blocks, in space, determines the emergent configurational properties of that space. For example, one consequence of the squaring law is that as we move objects from corner to edge and then to central locations in bounded spaces, total inter-visibility in the system decreases, as does visual integration (or universal visual distance) defined as how few visual steps we need to link all points to all others *Figure 11 (left)* The same applies to metric integration (or metric universal distance) defined as the sum of shortest paths between all pairs of points in the ambient space, which increases as we move the obstacle from corner to centre *(right)*.
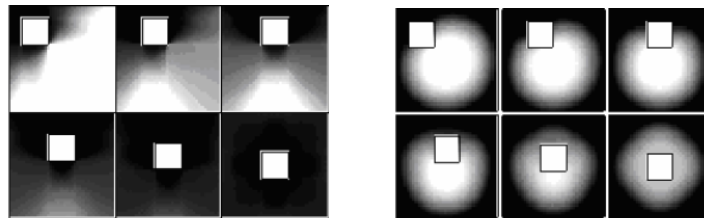

Figure 11 Moving an object from corner to centre decreases intervisibility (left – light means less visual distance to all other points, and dark more) and increases the mean length of trips (right – light is less metric distance, and dark more).

The same same squaring law governs the effect of shape (*Figure 6*): the more we elongate shapes, keeping area constant, the more we decrease inter-visibility and increase trip length in the ambient space. The effect of a long and short boundary is to create greater blockage in the system through the squaring law.
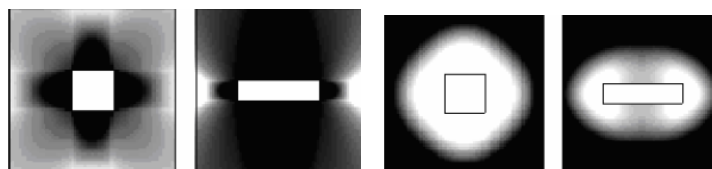

Figure 12 Changing the shape of an object from square to rectangular decreases inter-visibility and increase mean trip length. Again, light means less visual distance (left) and metric distance (right)

Even at this stage, this spatial law has a critical implication for cities: in terms of configurational metrics a short line and a long line are, other things being equal, metrically and visually more

efficient in linking the system together than two lines of equal length. (*Figure 7*), as would be a large space and a small space, compare to two equal spaces.
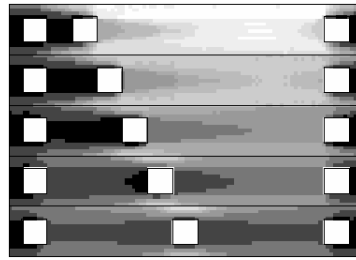

Figure 13 Other things being equal, a short and long line integrate more than two lines of equal length. Again, dark means less visual distance.

Another consequence is for the mean length of trip (or metric integration) from all points to all others in different types of grid, holding ground coverage of blocks, and therefore total travellable distance in the space, constant. In the four grids in *Figure 14*, darker (for clarity) means shorter mean trip length to all other points. Compared with the regular orthogonal grid (top left), interference in linearity on the right slightly increases mean trip length. But more strikingly, if we reduce the size of central blocks and compensate by increasing the size of peripheral blocks, we reduce mean trip length compared to the regular grid. This of course is the 'grid intensification' that we often note in looking at centres and sub-centres in cities. As so often, we find a mathematical principle underlying an empirical phenomenon. (Hillier 2000)
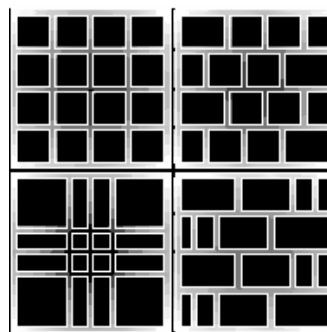

Figure 14 Changing the scaling of a grid changes mean trip length. In this case, for graphical clarity, dark means less metric distance from each point to all others. The mean distances for each system are: top left 2.53, top right 2.59, bottom right 2.71, bottom left 2.42

How we place and shape objects in space then determines the emergent configurational properties of that space. But what kind of block placing and shaping make space urban?
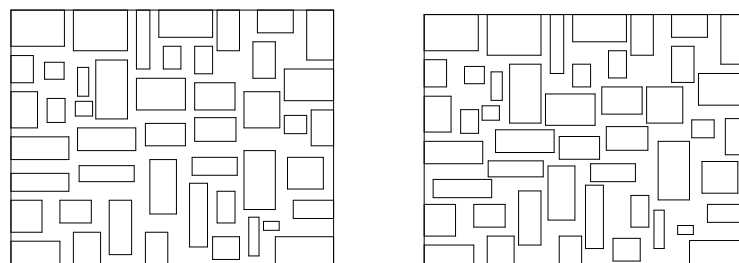

Figure 15 Two slightly different arrangements of identical blocks, with strong linear relations between spaces on the left and weak on the right

On the left of *Figure 15*, we aggregate buildings in an approximately urban way, with linear relations between spaces, so we can see where we are going as well as where we are. On the right

we retain the identical blocks but move them slightly to break linear connections between the spaces. If we then analyse metric and visual distances within the two complexes, we find that all to all metric distances (not shown) increases in the right hand case, so trips are on average longer, but the effect is slight compared to the effect on all to all visual distances, which changes dramatically (shown in *Figure 16*). Showing visual integration – dark mean less visual distance as before - we see that the left case identifies a kind of main street with side and back streets, so an urban type structure has emerged. But the right case has lost both structure and degree of inter-visibility. Even though the changes are minor, it feels like a labyrinth. We can see where we are but not where we might be.
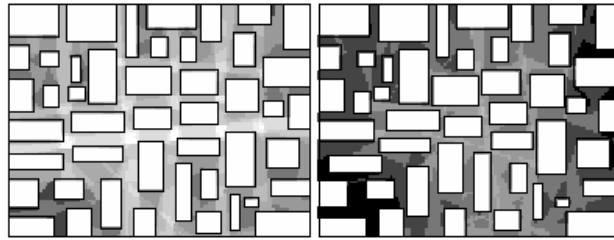

Figure 16 Visual integration analysis (light is high, and so low visual distances from all points to all others) showing how non-urban layout on the loses both integration and structure through the slight block changes

The effect on computer agents moving around the system is striking, if obvious. In *Figure 17* we move 10000 computer agents with forward vision in the space, again using the software by Alasdair Turner (Turner 2002). The agents randomly select a target within their field of vision, move 3 pixels in that direction, then stop and repeat the process. On the left, the traces of agent movement 'find' the structure of visual integration. On the right, they wander everywhere and tend to get trapped in fatter spaces. This is an effect purely of the configuration, since everything else is identical.
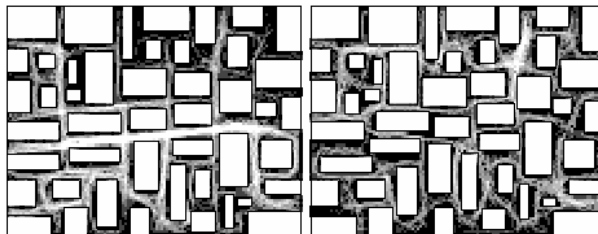

Figure 17 Traces of 10000 forward looking agents moving nearly randomly in two slightly different configurations. Light means many traces, dark few.

But what about human beings? Human beings do not of course move randomly, but purposefully, and successful navigation in an unfamiliar environment would seem to depend on how good a picture of the whole pattern we can get from seeing it from a succession of points within it. One way we might plausibly measure this property is by correlating the size of the visual field we can see from each point with the visual integration value (its visual distance from all others), so in effect measuring the relation between a *local* property that we can see from each point, and a *non-local* one that we cannot see (*Figure 18*)
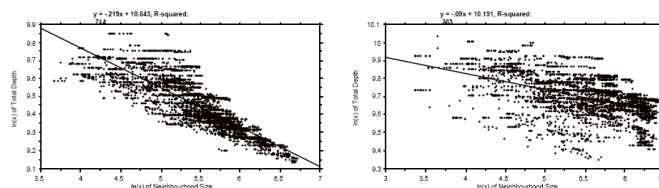

Figure 18 Intelligibility scattergams for the two layouts in Figure 15

In space syntax this is called this the *intelligibility* of the system. The r² for the 'intelligible' layout on the left is 0.714 while for the right case it is 0.267. Defined this way, the intelligibility of a spatial network depends almost entirely on its linear structure. Both field studies (Hillier et al 1987) and experiments (Conroy-Dalton 2001) suggest that this does work for humans. For example, Conroy Dalton took a linearised 'urban' type network (*Figure 19 left below*) and asked subjects to navigate in a 3D immersive world from left edge to 'town square' and back. As the traces show, they manage to find reasonable routes. But she then moved the (identical) blocks slightly to break the linear structure and reduce intelligibility (*Figure 19 right below*), and repeated the experiment. The subjects found the modified layout labyrinthine and many wandered all over the system trying to perform the same way-finding task.
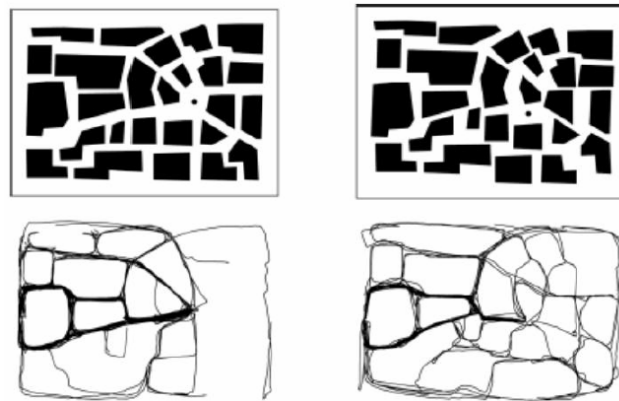

Figure 19 Trace of human agents navigating in an intelligible (left) and unintelligible (right) layout

So if, coming back to our aggregative process, we modify it by requiring those adding cells to the system to avoid blocking a longer local line if they can block a shorter one (*Figure 20*, left), we find a much more urban type layout emerges approximating the mix of long and short lines we find in real systems and emulating certain structural features (Hillier 2002). With the contrary rule — always block long lines (*Figure 20*, right) — we construct a labyrinth in which lines are of much more even length. So urban space networks seem to be shaped in some degree by a combination of spatial laws and human agency, with the human agents implementing, and so in a sense knowing, the spatial laws. The consistency we find in urban space patterns suggests that human beings 'know' the configurational laws of space in some sense – perhaps in the same sense that they 'know' simple 'intuitive physics' when they throw a ball of paper so that its parabola leads it to land in a waste paper basket.

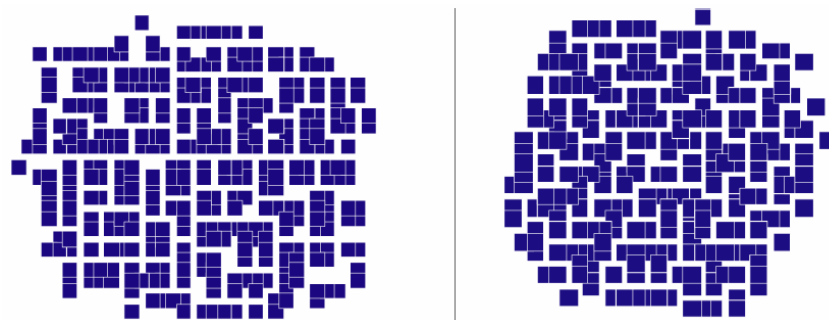*Figure 20 A layout generated by a 'conserve longer lines' rule (left) and one generated by the inverse rule*

Spatial agency

Spatial emergence is then governed by the *squaring law* though which the placing and shaping of objects in space creates emergent patterns, and this is why, simply to be intelligible to human beings, spatial networks must include enough long alignments, in proportion to the scale of the

settlement itself Hillier 2002) *Spatial agency* is then about the consequences of these emergent structures for the *functionality* of the system. As spatial emergence depends on a spatial law, so spatial agency depends on a spatio-functional law we call the law of natural movement: that other things being equal, the main determinant of movement rates in different parts of a network will be a function of the structure of the network itself.

To clarify this we may first reflect for on human movement. Spatially speaking, every human trip is made up of two elements: an origin-destination pair—every trip is from an origin space to a destination space—we can call this the *to-movement* component; and the spaces passed through on the way from origin to destination—we can call this the *through-movement* component. It is exactly these two elements of movement which are captured in the closeness (integration) and betweenness (choice) measures. Integration measures the accessibility of nodes as destinations from origins, then from the principle of distance decay (and other things being equal), we must statistically expect more movement potential for nodes that are closer to all others at some radius. Likewise, since choice measures the sequence of segments we pass through so we must expect a similar bias in real movement. In effect integration measures the to-movement, and choice the through-movement, potential of spaces and since we have used these to measure movement potentials of both kinds in urban networks, it would be surprising if these potential did not to some degree reflect real movement flows.

But this will depend on how people calculate distances in complex spatial networks, and this is a question, much discussed in the cognitive literature (for example Winter 2002, Timpf et al 1992, Hochmair & Frank 2002, Conroy-Dalton 2003, Duckham & Kulik 2003, Golledge 1995, Montello 1992, 1997, Sadalla 1980, Duckham, Kulik & Worboys 2003, Kim & Penn 2004) All three measure of distance used in DepthMap - shortest paths. fewest turns paths and least angle change have all been canvassed. But in (Hillier & Iida 2005) we suggest this can be resolved by correlating real flows with the spatial values produced in DepthMap by the three different definitions of distance. Accordingly, we applied the three weightings to the two measures of to and through movement potentials to make six different analyses of the same urban system, and correlated the resulting patterns of values for each segment with observed movement flows on that segment (Tables 1, 2), arguing that if across cases there were consistently better correlations

VEHICULAR MOVEMENT r² values for correlations between vehicular flows and shortest path, least angle and fewest turns analysis applied to accessibility and choice measures. Best correlations are marked *. Numbers in brackets indicate best radius in segments for accessibility measures.

	Gates	Measure	Least Length	Least angle	Fewest turns
BARNSBURY	116	accessibility choice	.131(60) .579	.678(90) .720*	.698(12) .558
CALTHORPE	63	accessibility choice	.095(93) .585	.837*(90) .773*	.819(69) .695
SOUTH KEN	87	accessibility choice	.175(93) .645	.688(24) .629	.741*(27) .649
BROMPTON	90	accessibility choice	.084(81) .475	.692*(33) .651*	.642(27) .588

PEDESTRIAN MOVEMENT r² values for correlations between pedestrian flows and shortest path, least angle and fewest turns analysis applied to accessibility and choice measures. Best correlations are marked *. 'a' or 'c' for combined multiple values indicates whether accessibility or choice is dominant. Numbers in brackets indicate best radius in segments for accessibility measures.

	Gates	Measure	Least length	Least angle	Fewest turns
BARNSBURY	117	accessibility choice	.119(57) .578	.719*(18) .705	.701(12) .566
CALTHORPE	63	accessibility choice	.061(102) .430	.637(39) .544*	.624*(36) .353
SOUTH KEN	87	accessibility choice	.152(87) .314	.523*(21) .457	.502(15) .526*
BROMPTON	90	accessibility choice	.111(81) .455	.623*(63) .513*	.578(63) .516

Tables 1 and 2 showing r² values for observed movement and spatial values

with one or other weighting, then the only logical explanation would be that this weighting reflects better how people are biasing spatial movement choices, since everything else about the system is identical. In fact, across four separate studies in areas of central London, we consistently found that geometric, or least angle weightings yields the strongest movement prediction, with an average of around 0.7 for vehicular movement and 0.6 for pedestrian, closely followed by the topological or fewest turns weighting. Metric shortest paths are markedly inferior in most cases, and in general, to-movement potentials are slightly stronger than through-movement potentials, though this varies from case to case. (Hillier & Iida 2005)

Once the law of natural movement is understood, it is clear that the link between the network configuration and movement flows is the key to the dynamics and evolution of the system. Because the network shapes movement, it also over time shapes land use patterns, in that movement-seeking land uses, such as retail, migrate to locations which the network has made movement-rich while others, such as residence, tend to stay at movement-poor locations. This creates multiplier and feedback effects through which the city acquires its universal dual form as a foreground network of linked centres and sub-centres at all scales set into a background network of residential space. Through its impact on movement, the network has set in train a self-organising processes by which collections of buildings become living cities.

Expanding this a little, we can say that there is a generic process of centre formation on something like the following lines. Every centre has a centre. It starts with a spatial seed, usually an intersection, but it can be a segment. The seed of a centre will have *destination* and *route* values at both local and global levels. Some - usually small - centres start because they are the focus of a local intensified grid – a local case – others because they are at an important intersection – a global case. Both global and local properties are relevant to how centres form and evolve. The spatial values of the seed for the centre will establish what we can call a *fading distance* from the seed which defines the distance from the seed up to which e.g. shops will be viable. This is a function of metric distance from the seed proportionate to the strength of the seed. The centre will grow beyond the fading distance established by the initial seed to the degree that further seeds appear within the fading distance, which reinforce the original seed. Again these can be local or global, and stronger or weaker. A centre becomes larger to the degree that it is reinforced by what are, in effect, new seeds created by the grid which allow the shopping to be continuous.

Centres then expand in two ways: linearly and convexly. Linear expansion, the most common case, will be along a single alignment or two intersecting alignments, and occurs when the reinforcers are more or less orthogonal or up to 45 degrees to the original alignment or alignments. Convex expansion will be when the shopping streets form a localised grid, and this occur when reinforcers occur on the parallel as well as the orthogonal alignment. So centres vary in the strength of their local and global properties and reinforcers, and the balance between them will tend to define the nature of the centre. Most centres will be in some sense strong in both in local and global terms, but differences in the balance between local and global will be influential in generating the scale and character of the centre. Centres also grow or fail through interaction with neighbouring centres at different scales, and some potential locations for centre fail to be realised due to the existence of centre close by, but the way in which the urban grid evolves tends to ensure that seeds for potential centres occur only at certain distances from each other (Hillier 2009.

The dual city of economic and social forces

We have then found our dual structure, and we can explain it. Within the envelope created by cognitive constraints – the need for the city to be intelligible in order to be usable – we can now see how economic and social forces put their different imprints on the city. The foreground structure, the network of linked centres, has emerged to maximise grid-induced movement, driven by micro-economic activity. Micro-economic activity takes a universal spatial form and this type of foreground pattern is a near-universal in self-organised cities. The residential background network is configured to restrain and structure movement in the image of a

particular culture, and so tends to be culturally idiosyncratic, often expressed through a different geometry which makes the city as a whole look spatially different. We call the first the *generative* use of space since it aims to generate co-presence and make new things happen, and the second *conservative* since it aims to use space to reinforce existing features of society. In effect, the dual structure has arisen through different effects of the same laws governing the emergence of grid structure and its functional effects. In the foreground space is more random, in the background more rule-governed, so with more conceptual intervention.
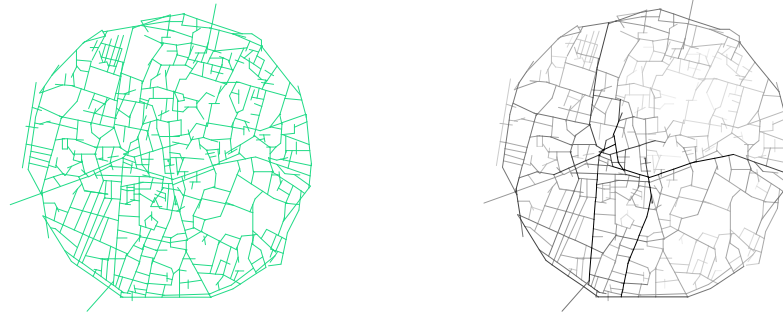

Figure 21 The old city of Nicosia (left) and its integration analysis, showing the deformed wheel core in spite of culturally differentiated residential space.

We can illustrate this most clearly in a city with more than one culture (now unfortunately separated): Nicosia (*Figure 21*). Top right is the Turkish quarter, bottom left the Greek quarter. Their line geometry is different. In the Turkish quarter, lines are shorter, their angles of incidence have a different range, and there is much less tendency for lines to pass through each other. Syntactically, the Turkish area is much less integrated than the Greek area. We can also show that it is less intelligible, and has less synergy between the local and global aspects of space. Yet in spite of these strong cultural differences in the tissue of space, we still find Nicosia as a whole is held together by a clear deformed wheel structure. This shows how micro-economic activity spatialises itself in a universal way to maximise movement and co-presence, while residence tends to be reflect the spatial dimension of a particular culture, and the expression is in the first instance geometrical. Since residence is most of what cities are, this 'cultural geometry' tends to dominate our spatial impressions of cities.

A meta-theoretical reflection: is this the way to the genetic code for cities

The dual *foreground-background* structure of the network of space in cities then reflects the differences between micro-economic and socio-cultural forces with each using the same underlying spatial and spatio-functional laws to achieve different effects. One of the difficulties of studying cities is that they seem to involve the interaction of physical, spatial, economic, social, cultural and cognitive processes, and in the past no models have existed for integrating such complex interactions.

Here, by studying the city in the first instance as what it seems to me, namely an aggregation of buildings creating a network of space, animated by movement and different kinds of activity, and bringing to light two simple laws, one governing the emergence of spatial patterns from the act of building, the other governing the impact of these emergent patterns on movement, we have put all these factors into a plausible relation to each other, and created a model in which each has its place, thought with the – useful – effect of discarding much disciplinary baggage on the way. The different aspects of the process of creating cities, which seem to have little to do with each other when viewed through disciplinary spectacles, all fall into place without at any point straining credibility.

It seems reasonable to advance the suggestion than that by expressing the complex processes of self-organisation through which cities come into existence as both spatial and functional systems, in terms of two simple., mathematically expressible laws, we are likely to be close to formulating the principles of a genetic code for cities. It is of course far from complete, and above all in need

of a general mathematical treatment. As we have said, space syntax is a unified theory of the city expressed as a patchwork of mathematical ideas. Even so, the theory has the merit of reflecting in quite a precise way the commonsense ways in which we experience and use cities. The model is close to urban reality rather than an abstraction from it.

A key feature of the theory, however, is the pervasive role it assigns to human cognition in the processes through which cities are created. In a sense it seems that the effect of human minds on both processes of both spatial emergence and spatial agency, is first to set the envelope of spatial possibility within which micro-economic and socio-cultural forces express themselves in space, and then to intervene within each of those processes to make them work in a particular way. In a sense then, the human mind is built into the very fabric of the city and its functioning. This to my mind is such a central theme in cities that it is the subject of another keynote paper I am giving this week at COSIT09. The paper is called *The city as a socio-technical system* and its central theme is precisely how human minds interact with the physical and spatial world so that the impact of human minds is pervasively present in the form of the city and its functioning.

bh
250809

References (to follow)

Carvalho, R., Penn, A. (2004) *Scaling and universality in the micro-structure of urban space*. Physica A 332 539–547

Conroy R (2000) *Spatial Navigation in Immersive Virtual Environments* PhD thesis University of London (UCL)

Conroy Dalton, R. (2003) *The secret is to follow your nose: route path selection and angularity*. Environment and Behavior 35 107–131

Duckham, M., Kulik, L. (2003) *Simplest paths: automated route selection for navigation*. In Kuhn, W., Worboys, M.F., Timpf, S., eds.: Spatial Information Theory:
Foundations of Geographic Information Science. Number 2825 in Lecture Notes in Computer Science. Springer-Verlag, Berlin 182–199

Duckham, M., Kulik, L., Worboys, M.F. (2003*) Imprecise navigation*. Geoinformatica 7 79–94

Golledge, R.G. (1995) *Path selection and route preference in human navigation: a progress report*. In Frank, A.U., Kuhn, W., eds.: Spatial Information Theory:
A Theoretical Basis for GIS. Number 988 in Lecture Notes in Computer Science. Springer-Verlag, Berlin 182–199

Hacking I (1983) Representing and Intervening Cambridge University Press

Hillier, B., Hanson, J.(1984) The Social Logic of Space. Cambridge University Press, Cambridge, UK (1984)

Hillier,**,** Burdett R, Peponis J, & Penn A (1987) - *Creating life: or, does architecture determine anything?* (with in Architecture & Behaviour, 3, 3, 233-250.

Hillier B et al (1993) - *Natural movement: or configuration and attraction in urban pedestrian movement -* Environment & Planning B: Planning & Design 19 20, 29-66

Hillier, B.(1996) Space is the Machine. Cambridge University Press, Cambridge, UK (1996)

Hillier, B. (1999) *The hidden geometry of deformed grids: or, why space syntax works, when it looks as though it shouldn't.* Environment and Planning B: Planning and Design 26(2) 169–191

Hillier B (2002) *A theory of the city as object* Urban Design International 7, 153-179

Hillier B & Penn A (2004) *Rejoinder to Carlo Ratti* Environment & Planning B: Planning and Design 31, 501-511

Hillier, B., Iida, S.(2005) *Network and psychological effects in urban movement.* In Cohn, A.G., Mark, D.M., eds.: Spatial Information Theory: COSIT 2005 in Lecture Notes in Computer Science 3693. Springer-Verlag, Berlin 475–490

Hillier B (2009) *Spatial sustainability: organic patterns and sustainable forms* Keynote paper to the Seventh Space Syntax Symposium, Stockholm

Hochmair, H., Frank, A.U. (2002) Influence of estimation errors on wayfinding-decisions in unknown street networks — analyzing the least-angle strategy. Spatial Cognition and Computation 2 283–313

Kim, Y.O., Penn, A. (2004) Linking the spatial syntax of cognitive maps to the spatial syntax of the environment. Environment and Behavior 36 483–504

Montello, D.R. (1992) The geometry of environmental knowledge. In Frank, A.U., Campari, I., Formentini, U., eds.: Theories and Methods of Spatial Reasoning in Geographic Space. Number 639 in Lecture Notes in Computer Science. Springer-Verlag,Berlin 136–152

Montello, D.R. (1997) The perception and cognition of environmental distance. In Hirtle, S.C., Frank, A.U., eds.: Spatial Information Theory: Theoretical Basis for GIS. Number 1329 in Lecture Notes in Computer Science. Springer-Verlag, Berlin 297–311

Sadalla, E.K. (1980) Burroughs, W.J., Staplin, L.J.: Reference points in spatial cognition. Journal of Experimental Psychology: Human Learning and Memory 6 516–528

Timpf, S., Volta, G.S., Pollock, D.W., Frank, A.U., Egenhofer, M.J. (1992) A conceptual model of wayfinding using multiple levels of abstraction. In Frank, A.U., Campari, I., Formentini, U., eds.: Theories and Methods of Spatial Reasoning in Geographic Space. Number 639 in Lecture Notes in Computer Science. Springer-Verlag, Berlin 348–367

Turner, A, 2002, '*Depthmap*, v2.11 (computer program) UCL', London, introduced in Turner, A, 2001,'*Depthmap: a program to perform visibility graph analysis*', Proceedings of the Third International Symposium on Space Syntax 2001, Atlanta, GA, pp. 31.1-31.9

Turner, A., Penn, A., Hillier, B. (2005) *An algorithmic definition of the axial map.* Environment and Planning B: Planning and Design 32(3) (2005) 425–444

Turner, A. (2001) Angular analysis. In: Proceedings of the Third International Space Syntax Symposium, Atlanta, GA, Georgia Institute of Technology 30.1–30.11

Wilson A (2000) Complex Spatial Systems: the modelling foundations of urban and regional analysis Prentice Hall

Winter, S. (2002) Modeling costs of turns in route planning. GeoInformatica 6
345–360