THE CITY AS A SOCIO-TECHNICAL SYSTEM: a spatial reformulation in the light of the levels problem and the parallel problem

Ketnote paper to the Conference on Spatial Information Theory, September 2009

Professor 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

On the face of it, cities as complex systems are made of (at least) two sub-systems: a physical sub-system, made up of buildings linked by streets, roads and infrastructure; and a human sub-system made up of movement, interaction and activity. As such, cities can be thought of as socio-technical systems. Any reasonable theory of urban complexity would need to link the social and technical sub-systems to each other. Historically, most urban models have sought to make the link through the concept of distance in the physical system as a cost in the social system, defining the physical sub-system as a set of discrete zones. Such models have proved practical tools, but over the years have contributed relatively little to the development of a more general theory of the city. Here we propose a more complex and, we believe, true-to-life model based on the definition of the physical sub-system of the city as a network of spaces – streets and roads - linking buildings, rather than as a system of discrete zones. This allows us to approach urban complexity in a new way.

Two key issues – theoretically perhaps the two key issues - in the study of complexity in general are the levels problem: how organised complexity at one level becomes elementary the next level up; and the parallel problem: how systems with different internal dynamics interact with each other. In (Cohen & Stewart 1993) a general framework for conceptualising these two problems is outlined: complex phenomena at one level commonly produce lawful (though rarely mathematically describable) emergent simplicities one level up which then have their own emergent dynamic, independent of the complex processes that created them. They call such emergent simplicities, in which 'chaos collapses', 'simplexities'. Simplexities of different kinds then interact and modify each other to create 'complicities', or complexes of simplexities, to construct the complexity of the real world. Here it is argued that this formulation captures the problem of complexity in cities as socio-technical systems, and that we need vertical theories to capture the relations across levels, and lateral theories to capture the relations of parallel systems. We outline a vertical and a lateral theory to account for generic aspects of the emergent complexity of cities.

However, both theories require an account of the linking mechanism, and here we show that since all actions that create cities are taken by human agents, the vertical and lateral linking mechanisms necessarily involve human minds, not in the sense of real historic individuals, but in the sense of a generalised individual acting according to spatial laws which are both objective and intuitively known, in the same sense that an individual who throws a ball of paper so that its parabola leads it to land in a waste paper basket intuitively 'knows' the law of physics. We call this generalised human subject the 'objective subject' of the city, and show that by virtue of being everywhere in space and time in the formation and working of the city, it everywhere imposes its point of view on it, so that cities are cognitive formations in an even more fundamental sense than they are socio-economic formations. The cognitive sets the envelope of possibility within which socio-economic processes create the city.

Vertical and lateral theories

On the face of it, cities as complex systems are made of (at least) two sub-systems: a physical sub-system, made up of buildings linked by streets, roads and infrastructure; and a human sub-system made up of movement, interaction and activity. As such, cities can be thought of as socio-technical systems. Any reasonable theory of urban complexity would need to link the social and technical sub-systems to each other. Historically, most urban models have sought to make the link through the concept of distance in the physical system as a cost in the social system, defining the physical sub-system as a set of discrete zones. Such models have proved practical tools, but over the years have contributed relatively little to the development of a more general theory of cities. Here we propose a more complex and, we believe, true-to-life model based on the definition of the physical sub-system of the city as a network of spaces – streets and roads -

linking buildings, rather than as a system of discrete zones. This allows us to approach urban complexity in a new way.

Two key issues – theoretically perhaps the two key issues - in the study of complexity in general are the *levels* problem: how organised complexity at one level becomes elementary the next level up; and the *parallel* problem: how systems with different internal dynamics interact with each other. In (Cohen & Stewart 1993) a general framework for conceptualising these two problems is outlined: complex phenomena at one level commonly produce lawful (though rarely mathematically describable) emergent simplicities one level up which then have their own emergent dynamic, independent of the complex processes that created them. They call such emergent simplicities, in which 'chaos collapses', 'simplexities'. Simplexities of different kinds then interact and modify each other to create 'complicities', or complexes of simplexities, to construct the complexity of the real world.

Here it is argued that this formulation captures well the problem of complexity in cities as socio-technical systems. To understand cities we need theories which deal with relations across levels and theories that deal with relations between parallel processes. We might call the former *vertical* theories, and the latter *lateral.* A *vertical* theory will be one which works across levels of emergent phenomena, showing how complex distributed processes produce emergent simplicities which then ignore the complexity of their creation and become independent forces at the emergent level in creating a further level of complexity. A *lateral* theory will then be one which shows how parallel but dynamically independent process, each with its own emergent simplicities, interact to shape each other.

Singularly clear examples of each can be found in the two most basic processes that create the city: the *vertical* emergence of a network of space from the process of aggregating buildings to create the physical city; and the *lateral* interaction of this emergent pattern with the processes by which different types of economic and social activity locate and organise themselves in space. The *vertical* process, we might say, creates the *spatial* city, the *lateral* process the *functional* city. The outcome of these vertical and lateral processes is the city which is at once a form of *spatial* patterning and at the same time a form of *social* and *economic* patterning.

These two processes, and the relations between them, are the subject of this paper. The ultimate focus is on a central question: what is the *mechanism* by which the vertical and lateral connections, on which the operation of the system depends, are made ? The answer proposed is the same in both cases: the *human mind*, using intuitive knowledge of *spatial* and *spatio-functional* laws, is the key mechanism. By this we mean that there exist spatial and spatio-functional laws which while being wholly objective and describable by simple mathematics, are intuited by people in the same sense that when we throw a ball of paper so that its parabola leads it to land in a waste paper basket we are intuiting the laws of physics. We do so in both cases because these are the laws of the world, and our bodies and minds then learn to operate under the constraints imposed by these laws. An implication of this is that we cannot understand the city without understanding the interaction between human minds and the form of the material world. This is why we must redefine the city as a socio-technical system if we are to develop effective theories about it.

Now this may seem at first a strange pair of ideas: that the emergence of the spatial *form* of the city and the emergence of its pattern of *functioning* both involve laws, and that these laws are somehow imposed on the city through *human cognition.* But both seem to be the unavoidable conclusions of the application of space syntax analysis to study cities over the past two decades. Here we show the research which has led us to these conclusions. First we show how the use of the space syntax methods to study cities has brought to light a remarkable series of spatial and spatio-functional *regularities* which are pretty well *invariant across cities*, suggesting that there is some kind of *universal city* underlying the diversity of real cities - a remarkable reflection if we bear in mind the heterogeneous social, economic and temporal circumstances in which the world's cities have been created. These regularities are for us the theoretical *explicandum* – that which is to be explained – of cities. Second, the use of space syntax as an *experimental tool* has brought to light simple *configurational* laws of space which deal precisely with the effect on emergent patterns of space of different kinds of physical intervention in it, such as placing blocks of buildings in space. We call this the law of *spatial emergence*, and have elsewhere called it the *law of centrality*. This is the law governing the *vertical* process through which the spatial form of the city emerges. Third, the use of space syntax as a tool to study *functional patterns* in cities has brought to light a fundamental relation between the *form* of the city and its *functioning*: that the network of space is in and of itself the primary influence on movement flows in different parts of the network, and through this, on emergent patterns of land use and densities, including the shaping of the pattern of centres and sub-centres. This is the law of *spatial agency*, and it governs the *lateral* process by which the form of the city shapes its functional patterns. We propose in effect that the *regularities* in cities – the *universal city* – can be explained by the *laws*, but that this happens in such a way as to be show that *human cognition* is the primary linking mechanism in both processes.

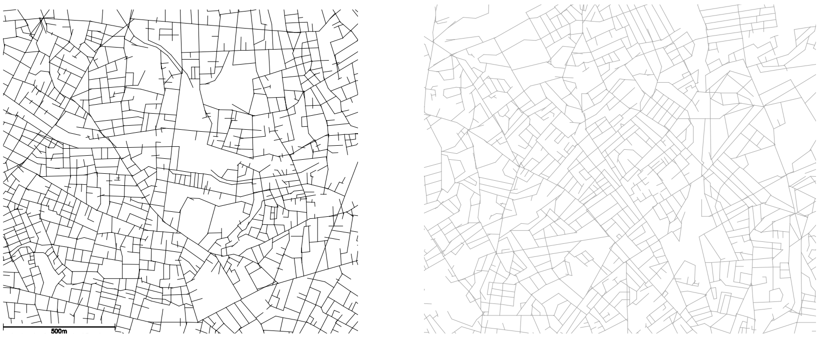
In what follows, we first show how space syntax has brought to light the regularities in the structures of the spatial networks of cities. Then we set out how the laws of the vertical and lateral processes came to light, and show how they have shaped the city into a spatio-functional whole. We then show how both processes can be seen in terms of this Cohen-Stewart model, and why this leads to the conclusion that the key mechanism in both the vertical and lateral processes is the human mind - not in the sense of real historic individuals, but in the sense of a *generalised individual* acting according to spatial laws which are both objective and intuitively known to humankind in general. We call this generalised human individual the *objective subject* of the city, meaning, in effect, *all of us*, and how we use our minds in making and use the city. By virtue of being everywhere in space and time in the formation and working of the city, the *objective subject* everywhere imposes its point of view on it, so that cities are cognitive formations in an even more fundamental sense than they are socio-economic formations. The cognitive sets the envelope of possibility within which socio-economic processes create the city. This is why we must revise our definition of the city as *socio-technical* system.

Regularities in urban spatial form

To bring to light the *regularities* in the spatial form of urban networks we must make us of the standard space syntax representation of the network, the *least line map*, or the smallest number of straight lines that cover the system while making all connections. 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).

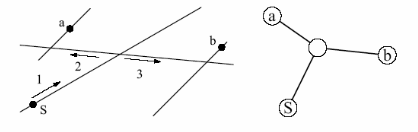
Examining least line maps for cities at all scales and in all parts of the world we find:

- that at all scales, from local areas to whole cities, cities 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.
- that in 'organic' cities (as defined above), the longer the line the more likely it is to be end-connected to another by a nearly straight connection (between about 5 and 25 degrees), creating sequences of such lines, which the eye instinctively identifies *Figure 1 and 2* when look at a least line map; the shorter the line the more likely it is to intersect with others at near right angles, creating local clusters of such lines. In geometrical cities, a similar pattern can be found but with straight rather than nearly straight long lines
- through these metric and geometric regularities, cities street networks acquire a *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.

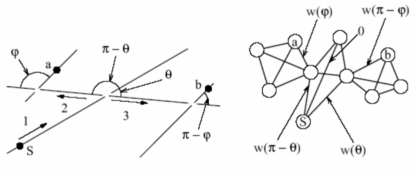


Figures 1and 2 Arbitrary sections of the least line maps of a section of Tokyo (left) and London (right)

Using the UCL Depthmap software, we can then bring to light *configurational* regularities in city structures. DepthMap breaks up the least line map into the street segments between junctions, allows 3 definitions of the distance between each segment and 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. 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*[1] 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* from each segment 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.



(a) line model (b) graph of (a)

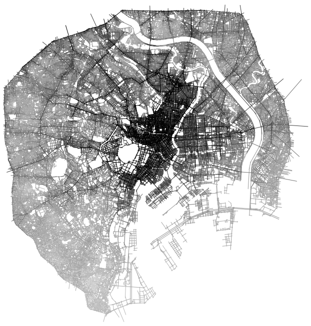


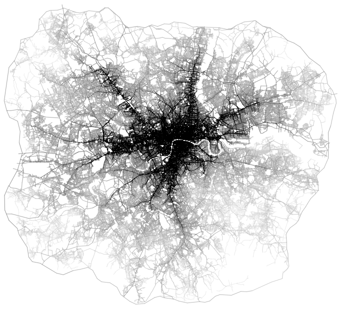
(c) segment model (d) graph of (c)

Figure 3 Line and segment representation of street networks and their graphs

Applying these measures to cities, we bring to light further regularities. For example:

- by colour banding mathematical values from red (dark) through orange and yellow to green and blue (light), meaning to strong to weak, we find in case after case, least angle integration (normalised closeness) 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 4* and *Figure 5*, for example, show the underlying deformed wheel pattern in both metropolitan Tokyo (with multiple rims) and London within the M25.





Figures 4 and 5 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

- Using the same colouring techniques, the least angle choice (betweenness) measure commonly identifies a network spread through the system, though strongest in the more syntactically central locations (see *Figure 5 and 6*).

Figures 6 and 7 Least angle choice (betweenness) analysis of London (left) and Tokyo (right) showing the network pattern in both cases

In other words, in spite of the differences in socio-economic and temporal circumstances in which cities grow, they seem to converge on common generic forms which have metric, geometric and configurational properties.

However the similarities between cities does not stop there. On close examination, for example:

- all cities seem to exhibit a property we call *pervasive centrality*, meaning that 'central' functions

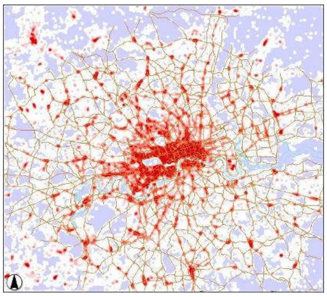


Figure 7 Mike Batty's map of the 168 main centres and sub-centres in London within the M25

> such as retail and catering concentrations diffuse throughout the network at all scales, from the city as a whole to the local network of streets. For example, *Figure 7* is Mike Batty's image of the 168 largest centres in London within the M25. By comparing *Figure 7* to *Figure 5* 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, *Figure 8* 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. 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)

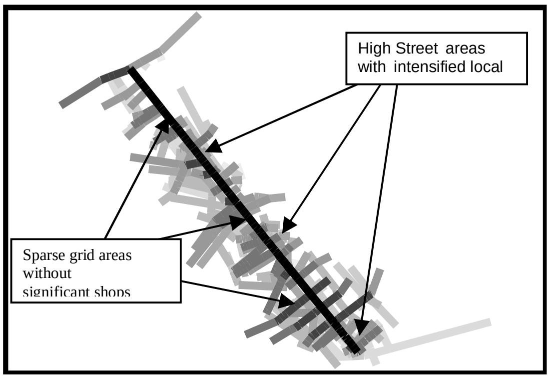


Figure 8 Grid intensification coincides with the high street areas in a London alignment

- if we then reduce the metric radius of the measures we then find the – much more numerous - smaller scale centres. For example, at radius 750 metres, all of the 'urban villages' in a section of north west London are picked out in red (dark). *Figure 9*

Figure 9 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

These effects are not confined to London or organic grids in general. The same kind of pattern of pervasive centrality was recently found in the historic grid based city of Suzhou in China. We have also show it to be the case in Brasilia. But it is critical that these effects are found in the least angle map and disappear if we substitute metric for least angle distance in the model. For example, Figure 10 (not shown - to be shown at conference) shows the pattern of shops in one of the unplanned areas of Jeddah in Saudi Arabia and Figure 11 (not shown - to be shown at conference) shows the least angle choice measure at a radius of 3.5 kilometres. The match between the two red patterns is remarkable. If we substitute metric for least angle distance Figure 12 (not shown - to be shown at conference) we find no relation to the functional patterns. The reasons are simple. In Figure 13, we consider three ways of diagonalising a grid. In the top case, the diagonal is regular and so the length of the diagonal route is identical to that of the right side peripheral route. Bottom left, we then create an upward kink on of the line elements, with the effect of marginally increasing the length of the diagonal route compared to the peripheral route. Bottom right, we create a downward kink on one line, so marginally shortening the diagonal route compared to the peripheral route, which we show following our usual colouring convention. It follows that with the most marginal changes of this kind, shortest routes will find complex diagonals or simple peripheral routes more or less arbitrarily. This is confirmed in the right figure where we construct a system in which the two diagonals compete, and movement shifts decisively to the downward link and so the shortest path route. Which route is selected by the shortest path algorithm will often then depend on very minor differences in angles, and so be virtually arbitrary.

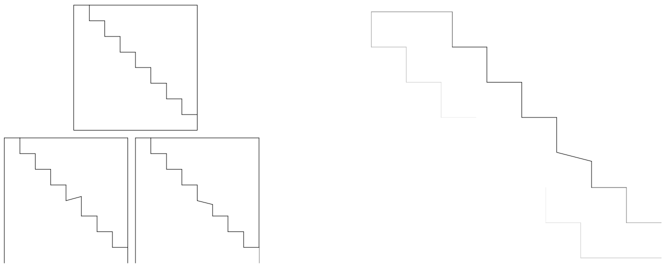


Figure 13 Different ways of diagonalising the grid, showing why minor geometrical changes can lead to near arbitrary changes in shortest paths

This arbitrary selection of complex diagonals as shortest paths will feature particularly strongly where a more regular grid system is associated with complex internal structures within grid islands. For example, in Beijing, shortest path choice analysis – right above - does not find the eight-lane boulevard between the Forbidden City and Tianamin Square, a boulevard which crosses Beijing east to west and is one of the busiest routes in Beijing. This is then a remarkable failure. In the case of Jeddah, least angle choice analysis without radius restriction (and so with reference to Jeddah as a whole) picks out the pattern of shops, though more weakly than with the local analysis, *Figure 14,* but substituting metric for least angle distance we find *Figure 15,* highlighting a nonsense route through the system, with innumerable changes of direction, and with no relation to the functional pattern. At best, we might say that metric analysis helps to identify taxi driver's routes !

Figure 14 Least angle choice (left) of an area of Jeddah and Figure 15 metric choice analysis (right), both at radius-n

A new definition of the city

The regularities that we find in cities with least angle analysis suggest a new definition of the city. Cities of all kinds, however they originate, 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. We suggest this is the proper generic definition of what a city is as a large object.

So what forces give the city this shape. We believe the answer lies 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 these laws 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. It is the *law of spatial emergence* which governs the *vertical* process through which the form of the city's spatial network emerges. The second phenomenon 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. The *law of spatial agency* governs the *lateral* process through which cities fit functional to spatial patterns.

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.

A vertical theory: spatial emergence as a law governed process

To understand the emergence of the spatial form of urban network – the vertical problem - 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.

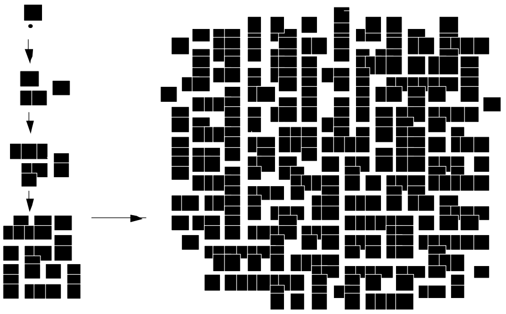


Figure 16 Aggregating dyads of open and closed cells by a restricted random process

If we take cell dyads (*Figure. 16*, 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?

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 17* 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.

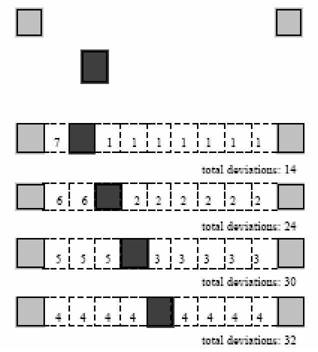
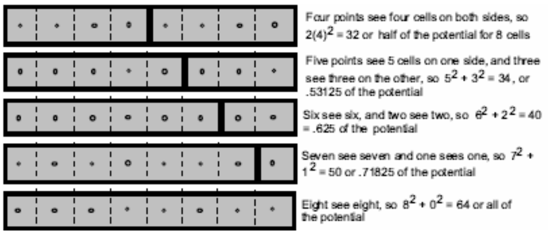


Figure 17 Moving an object between two others from edge to centre increases the sum of distances from all cells to all others

The same applies to intervisibility from all points to all others *Figure 18*. As we move a partition in a line of cells from centre to edge, the total inter-visibility from each cell to all others increases, though of course the total area remains constant.



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

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. 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:

$$2n^2 < (n - x)^2 + (n + x)^2 \tag{1}$$

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 19 (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).*

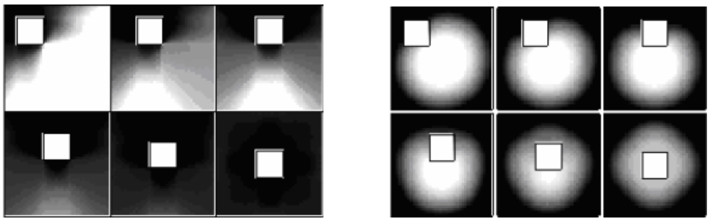


Figure 19 Moving an object from corner to centre decreases intervisibility (left – light means less visual distance to all other points, and dark more) and increases the mean length of trips (right – light is less metric distance, and dark more).

The same same squaring law governs the effect of shape *Figure 20*: 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. Even at this stage, this spatial law has a critical implication for

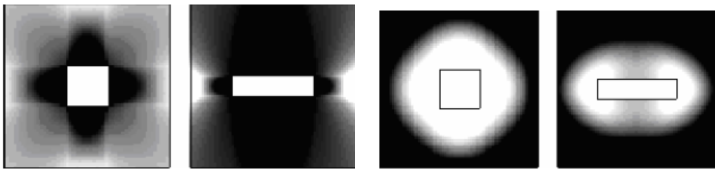


Figure 20 Changing the shape of an object from square to rectangular decreases inter-visibility and increase mean trip length. Again, light means less visual distance (left) and metric distance (right)

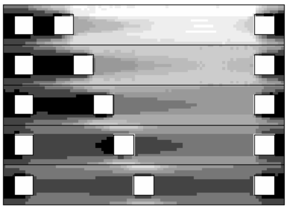


Figure 21 Other things being equal, a short and long line integrate more than two lines of equal length. Again, dark means less visual distance.

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 21*, as would be a large space and a small space, compared to two equal spaces.

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 22*, 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.

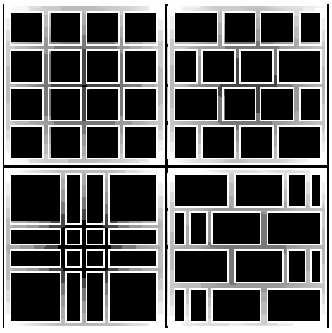


Figure 22 Changing the scaling of a grid changes mean trip length. In this case, for graphical clarity, dark means less metric distance from each point to all others. The mean distances for each system are: top left 2.53, top right 2.59, bottom right 2.71, bottom left 2.42

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?

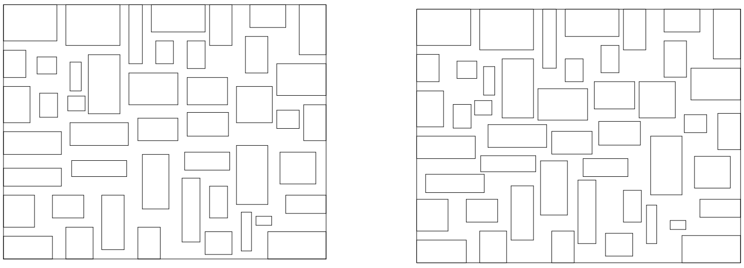


Figure 23 Two slightly different arrangements of identical blocks, with strong linear relations between spaces on the left and weak on the right

On the left of *Figure 23*, 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 24*). 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.

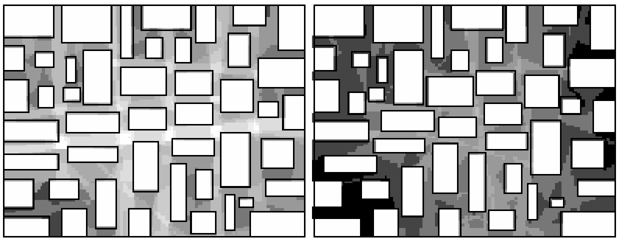


Figure 24 Visual integration analysis (light is high, and so low visual distances from all points to all others) showing how non-urban layout on the loses both integration and structure through the slight block changes

The effect on computer agents moving around the system is striking, if obvious. In *Figure 25* 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.

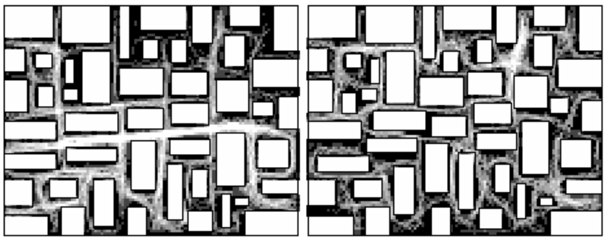


Figure 25 Traces of 10000 forward looking agents moving nearly randomly in two slightly different configurations. Light means many traces, dark few.

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 26*)

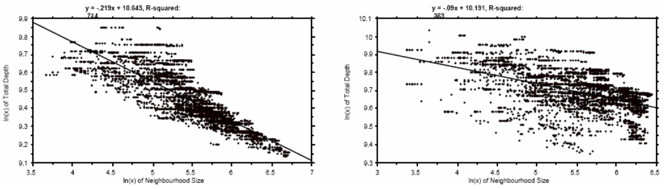


Figure 26 Intelligibility scattergams for the two layouts in Figure 15

In space syntax this is called this the *intelligibility* of the system. The r^2 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 27 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 27 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.

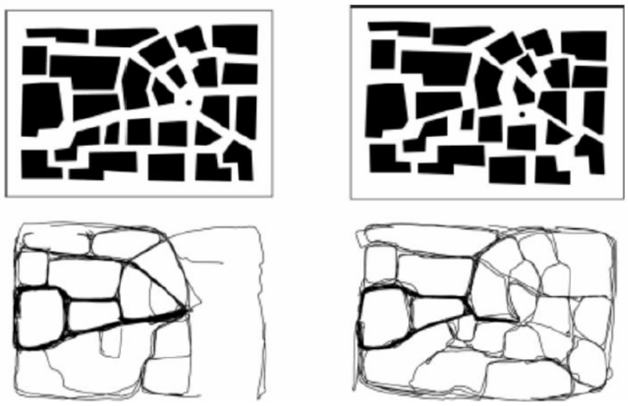


Figure 27 Trace of human agents navigating in an intelligible (left) and unintelligible (right) layout

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 28*, left) – we might call it a *preferential avoidance* rule ! - we find a layout emerges, which, while still not yet recognisably urban, approximates the mix of long and short lines we find in real systems (Hillier 2002). With the contrary rule — always block long lines (*Figure 28*, right) — we construct a labyrinth in which lines are of much more even length.

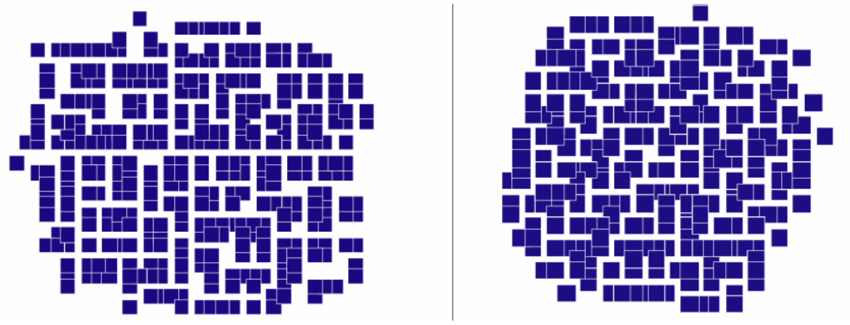


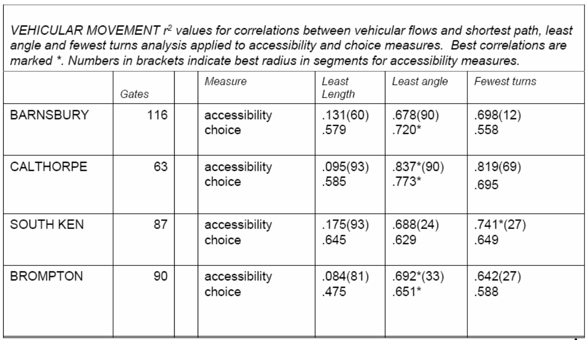
Figure 28 A layout generated by a 'conserve longer lines' rule (left) and one generated by the inverse rule

A lateral theory of spatial agency

The *vertical* process of *spatial emergence* is then shaped 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, and usable by, human beings, spatial networks must include enough long alignments, in proportion to the scale of the settlement itself Hillier 2002) The *lateral* process of *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* (Hillier et al 1993): 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.

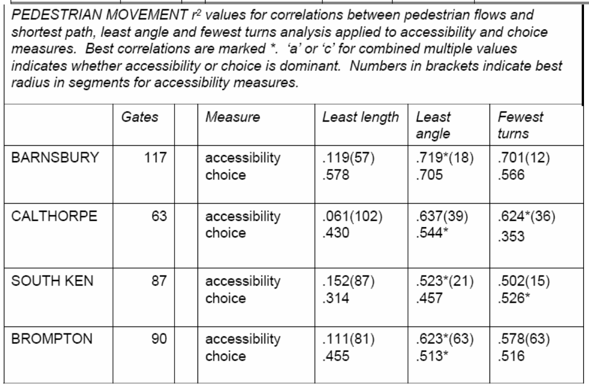
To clarify this we may first reflect for on human movement. Spatially speaking, every human trip is made up of two elements: an origin-destination pair—every trip is from an origin space to a destination space—we can call this the *to-movement* component; and the spaces passed through on the way from 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 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)



*VEHICULAR MOVEMENT r^2 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 r^2 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 r^2 values for observed movement and spatial values

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 *lateral* 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 the self-organising processes by which collections of buildings become living cities.

A key element of this will be the formation of centres and subcentres on something like the following lines. Every centre has a centre. Each centre 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.

The dual city of economic and social forces

Building on the *vertical* process, then, the *lateral* process is the means through which 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.

We can illustrate this most clearly in a city with more than one culture (now unfortunately separated): Nicosia *Figure 29*. 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.

Figure 29 The old city of Nicosia (left) and its integration analysis, showing the deformed wheel core in spite of culturally differentiated residential space.

The vertical and lateral processes as a simplexity-complicity duo

We see then that the form of the city and its characteristically urban functional patterns emerges from the *vertical* process of *form emergence* and the *lateral* process of *function emergence*. This is why in self-organised cities, things always seem to be in the right place. It is in the nature of the evolutionary processes through which cities acquire their spatial and functional form. But it is also clear that the vertical and lateral processes form a *simplexity-complicity duo* in the Cohen-Stewart sense. The low-level, step by step, aggregation of buildings, with the requirement that they be linked to continuous pattern of space, form an emergent pattern of space with its own independent structure – a *simplexity* - , and this independent structure, without regard for the complexities of its creation then shapes the spatial relation between economic and social processes, which have their own internal dynamics, but which become spatialised in the city through the movement law, so constituting a *complicity* formed by the interaction of socio-economic and spatial processes.

As we have seen, both the vertical, or simplexity, and the lateral, or complicity, processes are articulated through the intervening medium of spatial and spatio-functional laws. We see this in the fact that only a very small class of spatial forms are created by cities, a vanishingly small proportion of the possible forms that could be constructed with the same raw materials. This is all the more surprising if we bear in mind that almost all large scale random aggregates produced by the 'basic generative process' I described, are labyrinths, and labyrinths are nowhere found in cities, although we sometimes like to imagine that they are. On the contrary, cities tend towards an improbably high level of integration and intelligibility, and do so through the dual structure, which arises in the first instance from the distribution of line lengths. We can be in no doubt about this, because all cities exhibit dual structures related to the line length distribution, and in all cases the relation to functional processes is of the same kind. Cities are highly improbable forms, but of the same generic kind.

Since the astonishingly tight set of regularities that render cities non-labyrinthine are all expressions of the same basic laws of space, and all the actions that create cities are taken by human beings, it is hard to avoid the inference that the mechanism through which the laws of space reach the spatial form of the city is the human mind, and that this implies that human beings in general understand the laws of space. In fact careful observation of human behaviour easily shows that we do intuitively know the laws of space. Consider this. A group of people is 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? What Freddie knows is that by placing an object in the centre of a space, it creates more obstruction from all points to all others in that space than if the object is placed anywhere else. In this way, he seeks to draw attention to himself, so that people will interact with him rather than each other. In other words, at the age of two Freddie knows the laws of space and can use them to achieve social objective. *Figures 30, 31 and 32*

Or consider the politics of table shapes. If you take a simple shape, fill it with a fine grid and measure the distance from the centre point of each grid square to the centre point of all other grid squaresthe mean distance from central locations to all others is less than those of edge or corner locations. *Figure 33* We make the pattern clear by colouring low mean distances in red through to blue for high, using the same range in all cases. As we elongate the object, keeping area constant, we can see that mean distances increase, but the general centre to edge pattern is conserved. Now look more closely. If we look top right at the first elongation, we find that although the overall shape has higher mean distances, the mean distance in the centre of the long side is less than in the centre of the sides in the square shape. This is lost as we elongate more. So there is a certain point in the elongation of a square to a rectangular shape at which an optimal – in the sense of closer to all other points - edge location is created. It is this simple mathematical fact that is exploited in what we might call the politics of tables shapes, as in the next image. *Figure 34*

But this evidence of the intervention of human minds, knowledgeable of laws, in the processes of creating cities is circumstantial. We can find much more direct evidence for the intervention of minds in the lateral process through which the grid shapes movement and through this the overall functional patterns of the city. The fact that movement patterns reflect the objective distribution of least angle integration and choice in the system has clear implications. In choosing routes through the urban network, and so in all likelihood estimating distances, people must be using some kind of mental model of the urban grid involving geometric and topological elements, and since urban space can only be experienced as a set of discrete experiences, either of places or routes, these must then be synchronised into a larger scale pattern for this model to be formed. However, this is exactly the process described by cognitive sciences in moving from knowledge of routes to map-like knowledge (or 'survey' as they ineptly call it). So here we see that this kind of knowledge in human minds, by shaping movement patterns through the network, is shaping the emergent functional patterns in the city itself.

In other words, the key mechanism by which the vertical and lateral processes are linked is actually the human mind itself – in effect, all of us, taking decisions about how to move in the city. This in turn leads to another unavoidable conclusion: that because movement patterns reflect the large scale geometrical and topological structure of the network and not the local properties of space, the human mind is actually the means through which cities are created bottom up by the aggregation of building and spaces, but function top-down through the influence of the larger scale grid on movement patterns. And the mechanism by which this remarkable reversal takes place, at the moment when the vertical process first engages the lateral process, is the human mind itself. The vertical-lateral process which creates the city is then indecipherable without this knowledge of the intervention of human minds.

The objective subject

The proper theoretical conclusion of these explorations is, I believe, that the human cognitive subject is at its heart of the vertical and lateral processes that create the city, not simply in the sense of a series of real historical individuals located at specific points in time and space, but in terms of the invariance of the cognitive apparatus that those historical individual bring to the task of creating the city. We are talking in effect of a generalised individual located at all points in time and space in the city and everywhere imposing its cognitive apparatus on the ambient city. We might call this generalised individual the *objective subject* of the city creating process, and therefore of the city.

If, then, it is the case that the city has an objective subject which plays a critical role both in the 'vertical' form-creating process by which the accumulation of built forms creates an emergent spatial pattern, and in the 'lateral' form-function processes by which the emergent spatial pattern shapes movement and sets off the process by which an aggregate of buildings becomes a living city, then what does this imply for our paradigms of the city ? The field is broadly split between the social physics paradigm, which seeks to understand the formation of the physical city as the product of spatialised economic processes, and the humanistic or phenomenological paradigm which seeks to understand the city through our direct experience of it. The social physics view is essentially a mathematical view of the city, while the phenomenological view more or less precludes mathematics. The effect is to create paradigms which are as irreconcilable methodologically as they are theoretically. The split is made to appear natural by the way we conceptualise our field meta-theoretically as being about the relations between environments simply as material objects and human beings as experiencing 'subjects'.

If the argument in this paper corresponds in any sense to what really happens in cities, then it is clear that environmental 'objects' and human 'subjects' are deeply entangled with each other, with the 'subjective' appearing in the 'objective' world as much as the objective world appears in the human subject. Nor is it the case that the object side of the urban system can be dealt with mathematically and the subject side only qualitatively. The fact that the city is shaped by the human cognitive subject does not lessen its mathematical content since the cognitive processes by which the subject intervenes reflect mathematical laws (Hillier 2005, 2007). We cannot understand the generation of the material form of the city without understanding the formal aspects of the cognitive subject's role in shaping the city, nor understand the experience of the city without knowledge of the formal shape the city acquires under the influence of cognitive subjects.

It follows that we cannot progress while the paradigm split remains. Space syntax was originally created to try to find links between the two previously irreconcilable domains of the city, the city of people and the city of things, hence the 'social' logic of space. The project for space syntax research must now be to engage with the problematics of both the mathematical and humanistic paradigms in the hope and expectation that by finding how each is present in the other we will progress towards synthesis (Hillier 2005). It could also be instructive for the study of human-mediated complex systems in general, not excluding society itself (see Hillier 2007)

References

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

Cohen J & Stewart I (1994) *The Collapse of Chaos* Viking Penguin

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

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

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

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