Cities as movement economies

Bill Hillier

Unit for Architectural Studies, The Bartlett School of Graduate Studies, University College London, Gower Street, London WC1E 6BT, UK

This paper is taken from the forthcoming book, *Space is the Machine* (Cambridge University Press, 1996) which brings together some of the recent developments in applying configurational analysis to issues of architectural and urban theory. The paper reports a fundamental research finding: that movement in the urban grid is, all other things being equal, generated by the configuration of the grid itself. This finding allows completely new insights into the structure of urban grids, and the way these structures relate to urban function. The relation between grid and movement in fact underlies many other aspects of urban form: the distribution of land uses such as retail and residence, spatial patterning of crime, the evolution of different densities and even the part–whole structure of cities. The influence of the fundamental grid–movement relation is so pervasive that cities are conceptualized here as 'movement economies', in which the structuring of movement by the grid leads, through multiplier effects, to dense patterns of mixed use encounter that characterize the spatially successful city.

The physical city and the functional city

It is a truism to say that how we design cities depends on how we understand them. In the late twentieth century, this truism has a disquieting force. Cities are the largest and most complex artefacts that humankind makes. We have learned long and hard lessons about how we can damage them by insensitive interventions. But the growth of knowledge limps painfully along through a process of trial and error in which the slow timescale of our efforts, and the even slower timescale of our understanding, make it almost impossible to maintain the continuity of experience and study which we might hope, in time, would give rise to a deeper, more theoretical understanding of cities.

Even so, a deeper theoretical understanding is what we need. We are at a juncture where fundamental questions about the future of our cities – should settlements be dense or sparse, nucleated or dispersed, monocentric or polycentric, or a mix of all types? – have been raised by the issue of sustainability.[1] It is widely acknowledged that to make cities sustainable we must base decisions about them on a more secure understanding of them than we have now. What is unclear is what we mean by a better understanding. Physically, cities are stocks of buildings linked by space and infrastructure. Functionally, they support economic, social, cultural and environmental processes. In effect, they are means–ends systems in which the means are physical and the ends functional. Our most critical area of ignorance is about the relation of means to ends, that is of the physical city to the functional city. The fact that sustainability is about ends and the controls largely about means, has exposed our ignorance in this critical area.

One reason for this ignorance is the compartmentalization that has developed over the past quarter century among the disciplines concerned with the city. There is now a deep split between those who are preoccupied with analysis and control of the social and economic processes which animate the city, and who for the most part call themselves planners, and those concerned with physical and spatial synthesis in the city, who call themselves urban designers. This split is now, in effect, a split between understanding and design, between thought and action.

From the point of view of our ability to act on the city, there are two consequences. The first is a form–function gap: those who analyse urban function cannot conceptualize design, while those who can conceptualize design guess about function. The second is a scale gap. Planning begins with the region, deals reasonably with the 'functional city', that is the city and its 'dependences' (as the French say of outlying buildings) but barely gets to the urban area in which we live. Urban design begins with a group of buildings, gets to the urban area, but hesitates at the whole city for fear of repeating the errors of the past when whole city design meant over-orderly towns which never quite became places. Neither applies itself to our need to understand the city as a spatial and functional whole.

One effect of this disciplinary apartheid has been a complete failure to come to terms conceptually with what seems at first to be the simplest thing about the city: the fact that it is a large, apparently complex physical and spatial object, one which is at once a record of the functional processes which historically created it, and at the same time the strongest constraint on future development. Most attempts to use computers to model the ways in which cities work, for example, have dealt with the physical aspects of the city only at the grossest level, far above the level at which most interventions are made. Since the aim of an urban model is to try to bring the structural and dynamic complexities of cities as means–ends systems within the scope of reasoned decision-making about physical and spatial interventions, this has been a critical weakness.[2]

The fact that the physical city has proved most difficult to model effectively is probably due to two things. First, the physical and spatial structure of cities appears, for the most part, to be the rather disorderly outcome of a long history of small-scale, incremental changes which accumulate over time to produce patterns with neither geometrical nor functional simplicity. Until recently, the types of pattern that result from these quasi-organic processes have not seemed tractable to any obvious method of analysis. Consequently they were neglected. Second, the incremental ways in which economic and social processes create the city's physical and spatial patterns seem in themselves to be quite complex, involving feedback and multiplier effects, and interaction between different scales. Processes of urban growth and change seem to exhibit both 'emergence', by which unforeseen macro changes result from a series of micro changes, and the contrary effect, by which macro changes produce unforeseen effects at the micro scale. Again, until recently, there have not been obvious ways of modelling such processes.

The apparent intractability of the city as a physical and spatial object afflicts the synthesists as much as the analysts. If we look to urban designers for an analysis of the object of their design attention, we find much moral earnestness about such matters as the creation of 'places' as rich and complex as those found in traditional cities, but little analytical endeavour to understand how the physical and functional cities of the past gave rise to such 'places'. The current preoccupation with 'place' seems no more than the most recent version of the urban designer's preference for the local and apparently tractable at the expense of the global and intractable in cities. However, both practical experience and research suggest that the preoccupation with local place gets priorities in the wrong order. Places are not local things. They are moments in large-scale things, the large-scale things we call cities. Places do not make cities. It is cities that make places. The distinction is vital. We cannot make places without understanding cities. Once again we find ourselves needing, above all, an understanding of the city as a functioning physical and spatial object.

Multifunctionality and the part–whole problem

Where should we then find a starting point for an enquiry into the form and functioning of cities, in the hope of founding a theory of cities as means–ends systems? In situations where new theories are needed, there is a useful rule. At every stage in the development of our understanding of phenomena, we already have in our minds some conceptual scheme through which we interpret and interrelate the phenomena that we see.[3] Usually there are irritating anomalies and problems at the edges of these conceptual schemes. The rule is that instead of keeping these problems at the edge of our field of vision, and accepting them as anomalies, we should bring them centre stage and make them our starting point. We should, in effect, start from what we cannot explain rather than what we think we can.

There are two such great anomalies in our current ways of seeing cities. The first is the problem of multifunctionality. Every aspect of the spatial and physical configuration of the city form seems to have to work in many different ways – climatically, economically, socially, aesthetically, and so on – with the additional difficulty that form changes only slowly while function changes rapidly. The second is the part–whole problem, or as some might prefer, the place–city problem, that is, the fact that in most cities made up of parts with a strong sense of local place it is almost impossible to make a clear morphological distinction between one part and another, at least not at the level at which it could inform design.

If the theory set out here is anywhere near right, then it will become clear that these two issues are rather more than closely related: they really are the same problem, because all functions relate to the form of the city through two generic functional factors – how we as individuals find the city intelligible, and how we move around in it. These generic factors are so powerful that all other aspects of function pass through them and influence the urban form through them. This is so because in cities, as in buildings, the relationship between form and function passes through space. How we organize space into configuration is the key both to the forms of the city, and how human beings function in cities.

The theory to be set out here is based on one central proposition: that the fundamental correlate of the spatial configuration is movement. This is the case both in terms of the determination of spatial form, in that movement largely dictates the configuring of space in the city, and in terms of the effects of spatial form, in that movement is largely determined by spatial configuration. The principal generator of the theory set out here is the discovery, through recent research, that the structure of the urban grid considered purely as a spatial configuration, is itself the most powerful single determinant of urban movement, both pedestrian and vehicular. Because this relation is fundamental and lawful, it has already been a powerful force in shaping our historically evolved cities, by its effect on land use patterns, building densities, the mixing of uses in urban areas and the part–whole structure of the city.[4]

The result now available suggests that socio-economic forces shape the city primarily through the relations between movement and the structure of the urban grid. Well functioning cities can therefore, it will be suggested, be thought of as 'movement economies'. That is, it is the reciprocal effects of space and movement on each other (and not, for example, aesthetic or symbolic intentions) and the multiplier effects on both that arise from patterns of land use and building densities, which are themselves influenced by the space–movement relation, that give cities their characteristic structures, and give rise to the sense that everything is working together to create the special kinds of wellbeing and excitement that we associate with cities at their best.

It will be suggested as a consequence of these arguments that our view of the city in the recent past has been afflicted by conceptions of space which are at once too static and too localized. We need to replace these by concepts which are dynamic and global. Both can be achieved through the configurational modelling of space, using the power it gives us both to capture the complexities of urban form, and bring these analyses to bear on design.

Form and function in space are not independent

We must begin by making a few basic observations about space and its relation to function. We tend to think of the form and function of space as two quite independent things. Space is a shape, and function is what we do in it. Set up this way, it is hard to see why there should be any relation between the two, and even hard to see how any relation could be a necessary one.

But if we think a little more carefully about how human beings operate in space, we find everywhere a kind of natural geometry to what people do in space. Consider, for example, Fig. 1. At the most elementary level, people move in lines, and tend to approximate lines in more complex routes, as in the first figure. Then if an individual stops to talk to a group of people, the group will collectively define a space in which all the people the first person can see can see each other, and this is a mathematical definition of convexity in space, except that a mathematician would say points rather than people. The more complex shape of the third figure defines all the points in space, and therefore the potential people, that can be seen by any of the people in the convex space who can also see each other. We call this type of irregular, but well defined, shape a 'convex isovist'. Such shapes vary as we move about in cities, and therefore define a key aspect of our spatial experience of them.

There are relationships, then, between the formal describability of space and how people use it. These elementary relationships between the form of space and its use suggest that the proper way to formulate the relation is to say that space is given to us as a set of potentials, and that we exploit these potentials as individuals and collectivities in using space. It is this that makes the relation between space and function analysable, and to some extent predictable. By dividing up urban space, which is necessarily continuous, in different formal ways we are

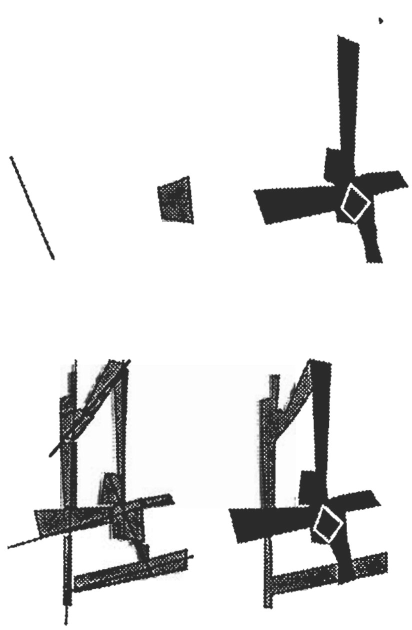


Fig. 1.

likely to be dividing it up according to some aspect of how human beings function.

Consider, for example, Fig. 2a which is the plan of Rome, in which the customary representation with the buildings in black and the space white has been reversed to draw attention to the fact that it is the black structure of space that is our focus of concern.[5] Figure 2b is then one possible structure within Fig. 2, the fewest and longest lines that cover the open space of Rome, and therefore form its potential route matrix. Figure 2c is another such structure: all the convex elements we call public open spaces together with their isovists. By definition, this includes all the lines that pass through the spaces and relate them in the urban structure as a whole. Note how they link up to form global clusters. We immediately see how mistaken we would be to see Roman squares as local elements. The isovists show they also form a global pattern.

All these ways of looking at space can be seen as layers of spatial structuring, coexisting within the same plan, each with its own contribution to intelligibility and function. A spatial layout can thus be seen as offering different functional potentials. What is it like to move around in it? Does it have potential to generate interaction? Can strangers understand it? And so on. All these questions are about the relationship of space as formal potentials to different aspects of function. A layout can thus be represented as a different kind of spatial system according to what aspects of function we are interested in.

The shape of space in the City of London

Let us now look in more detail at a case that is much closer to home: the City of London, for no better reason than that it has been as often criticized as 'haphazard' as praised as 'organic' – but never explained properly. The plan of the 'square mile' (in fact it is neither square nor a mile) is shown in Fig. 3a using the black on white convention to emphasize that it is space we are looking at. Figure 3b homes in on one of the allegedly 'labyrinthian' back areas of the City between Cornhill and Lombard Street, taken from the Rocque map of 1746. We say allegedly because although it looks so in plan, it does not seem in the least labyrinthian to the person

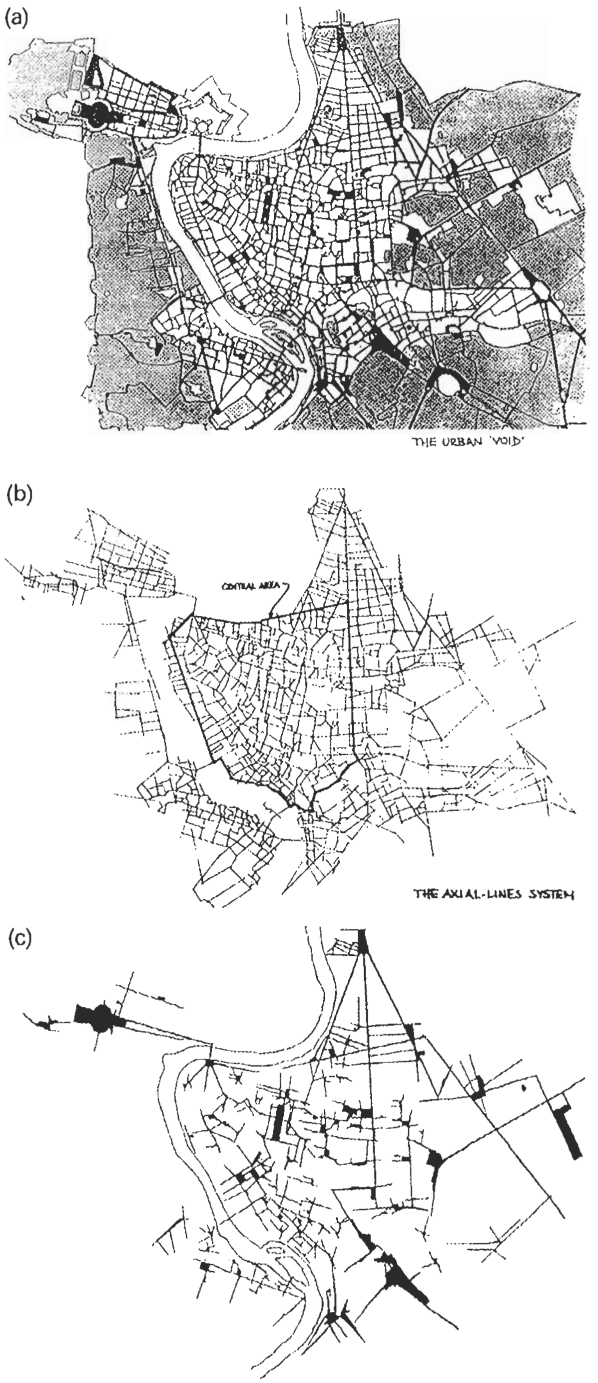


Fig. 2. (a) Plan of Rome, Italy; (b) Axial map of Rome, Italy; (c) Public open spaces in Rome.

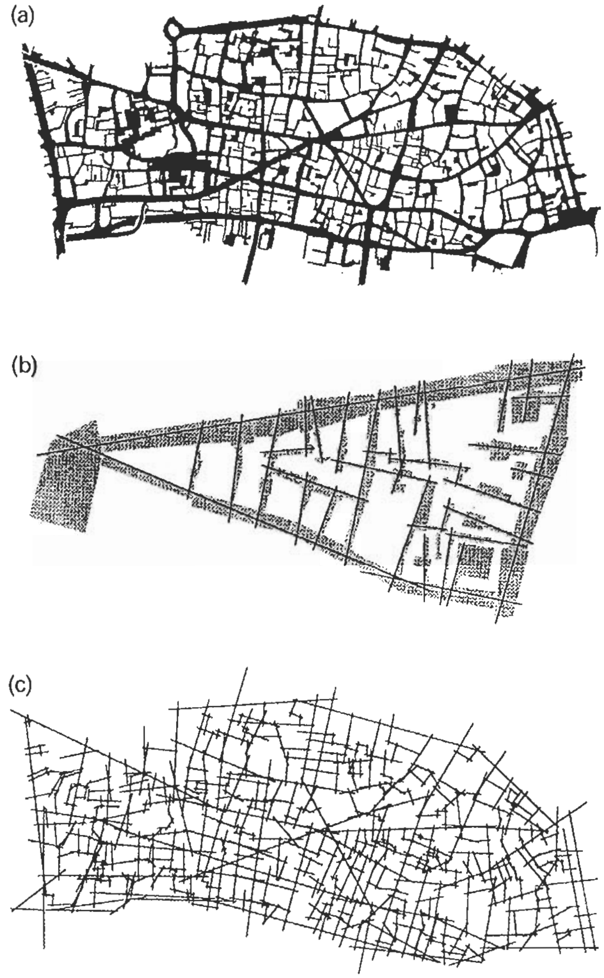


Fig. 3. (a) Black and white illustration of the public open space of the City of London as it is today; (b) Close-up of the one and two dimensional space structure of the area between Cornhill and Lombard Street in 1677; (c) Axial map of the City of London as it is today.

moving at ground level. On the contrary, it seems highly intelligible. How does this happen? The technique is simple. The space structure is admittedly highly broken up into 'convex' spaces – but there are always lines which link the convex spaces together, usually several at a time. Sometimes the line 'just about' gets through the spaces formed by the buildings, sometimes more easily. But because people move in lines, and need to understand lines in order to know where they can go, this means that the space structure is easily intelligible from the point of view of movement.

In fact, the pattern is slightly subtler. There is for the most part a 'two-line logic' in that if you pass down a line that you can see from the main grid, the next line will take you either out of the back area again to some significant spatial event – say a larger piece of space or a significant building – within the back area. This means that wherever you go, there is usually a point from which you can see where you have come from and where your next point of aim might be. This is the opposite of labyrinthian. As observation will confirm, the effect of this spatial technique is that the back areas become normally and naturally used for movement as part of the urban space pattern. There is no inhibition or sense of territorial intrusion in these areas.

This two-line logic is not the only constant property of these small-scale complexes. We also find that nearly every convex element, including the narrow ones that enter the back areas, as well as the fatter ones we find within the areas, has building entrances opening onto it. In the city, a fascinating cultural practice has augmented this: even in inclement weather, doors to buildings tend to be left open, often showing to the outside world one-way up stairs or down and another into the ground-level premises.

The effect of these apparent rules about how buildings relate to open space is to create two 'interfaces'. First, there is a close relation between those within the building, and those outside. Second, there is a natural mingling between those who are using the space outside the buildings, and those who are passing through. There is no sense of lack of privacy or intrusion. Nor is there any pressure to interact, though this is available if required. All we have is a relation of copresence between groups doing different things. Such copresence seems unforced, even relaxed. It is the product of a two-way relation from the convex spatial element: one into the building, the other to the larger scale through the line structure. The larger and smaller scales of space are held together by this spatial technique.

Now let us zoom out to the larger scale. Figure 3c is an 'axial map' of the city as a whole, that is, the least set of straight lines that pass through all the open space in Fig. 3a. The first thing we see when looking at the larger scale – that is at the longer lines – is that the tendency of lines 'just about' to pass through convex space is still there. It is just possible, in spite of the sinuous curves of the buildings, to see down Lombard Street from one end to the other, and it is just about possible to see from the Bank interchange through the whole of Cornhill into Leadenhall Street as far as Billiter Street. In both cases the line ends by striking the facade of a building at a very open angle, and from this it seems natural to infer continuation of potential movement in that general direction.

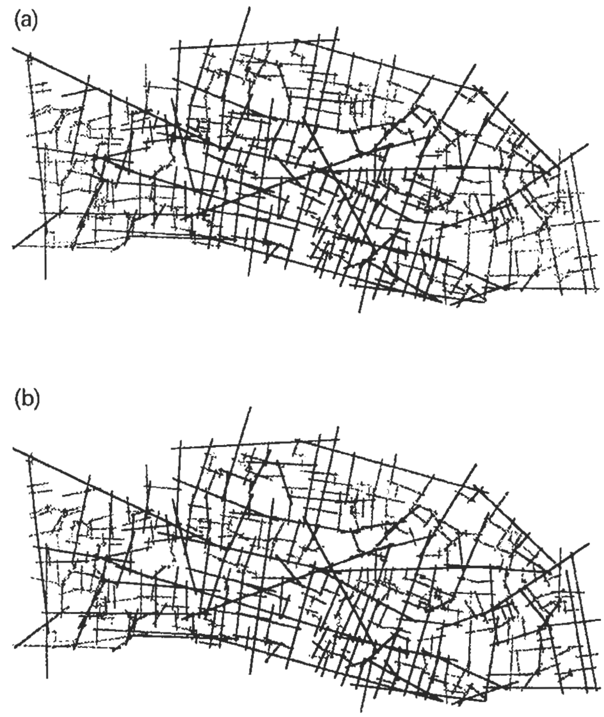
These improbably extended 'just about' lines create another effect which one must search a little to find, and perhaps go back to the old map to verify. It is that if one enters any of the old City gates and proceeds following only a rule that requires you to take the longest line available at any time (without going back on yourself) then in each case from somewhere on the second line a line opens up from which the Bank interchange (the old centre of the City) can be seen. Again, we find a simple two-line logic underlying apparent complexity, and again we need have no doubt about its functional implication. It accesses the stranger to the heart of the city. An automaton could find the centre – so a stranger could.

However, when we compare the two levels at which we find this two-line logic, there is a geometric difference which we can summarize in a simple principle: the longer the line the more likely it is to strike a building facade at an open angle; the shorter the line, the more likely it is to strike a building at a right angle. This is exactly the opposite of the current rather pompous urban fashion to end major axes at right angles on major building facades. Historically this usually occurs where urban space is taken over for the symbolic expression of power, whereas the City's urban space structure is about the movement required to create a dense encounter field. The right angle relation of facade to line is used in the City, as it were, to illuminate the smaller-scale and spatially more complex areas, and to make them visible from the larger-scale grid. Thus we begin to see not only that there is an interior logic to the city's apparently disorderly grid, but that this inner logic is fundamentally about movement, and the potential that movement gives for creating copresence. We see that many of the properties of urban space that we value aesthetically are a product of this functional shaping of space.

These consistencies in spatial patterning show how the City is put together locally, and how it therefore works as a series of experiences. But the city also acquires a global form. To understand this, and why it is important, we must begin to formalize our understanding a little. It will turn out that the line pattern of the city is the most important to its global structure, and we must therefore begin by examining this if we wish to move the focus of our analysis from the local to the global. We may begin by a simple observation: that to go from any line to any other one must pass through a certain number of intervening lines (unless of course the origin line directly intersects the destination line). Each line thus has a certain minimum line 'depth' from another, which is not necessarily a function of distance. It follows that each line has a minimum average line 'depth' from all other lines in the system. Because lines will always be shallow from some lines and deep from others, one might expect that this would average itself out. The surprising thing is that it does not. There are substantial differences in the mean depth of lines from all others, and it is these differences that govern the influence of the grid on movement in the system: roughly, the less depth to all other lines, the more movement; and the more depth the less movement.

These configurational pictures of the City from the point of view of its constituent lines can be measured exactly through the measure of 'integration'. The 'integration value' of each line reflects its mean linear 'depth' from all other lines in the system. We can then map these integration values, and produce a global integration map of the whole of a city, as in Fig. 4a. We can also produce another highly informative map, one in which we calculate integration only up to three lines away from each line in every direction, and which we therefore call 'local integration', or radius-3 integration, in contrast to 'global' or radius-n integration (Fig. 4b).

Integration values in line maps are of great importance in understanding how urban systems function because it turns out that how much movement passes down each line is very strongly influenced by its 'integration value' calculated in this way, that is by how the line is positioned with respect to the system as a whole.[6] In fact, it is slightly more subtle and depends on the typical length of journeys. Pedestrian densities on lines in local areas can usually be best predicted by calculating integration for the system of lines up to three lines away from each line (radius-3 integration), while cars on larger-scale routes (though not in local areas, where radius-3 is the best predictor) depend on higher radius integration because car journeys are on the whole longer and motorists therefore



Figs 4a & b.

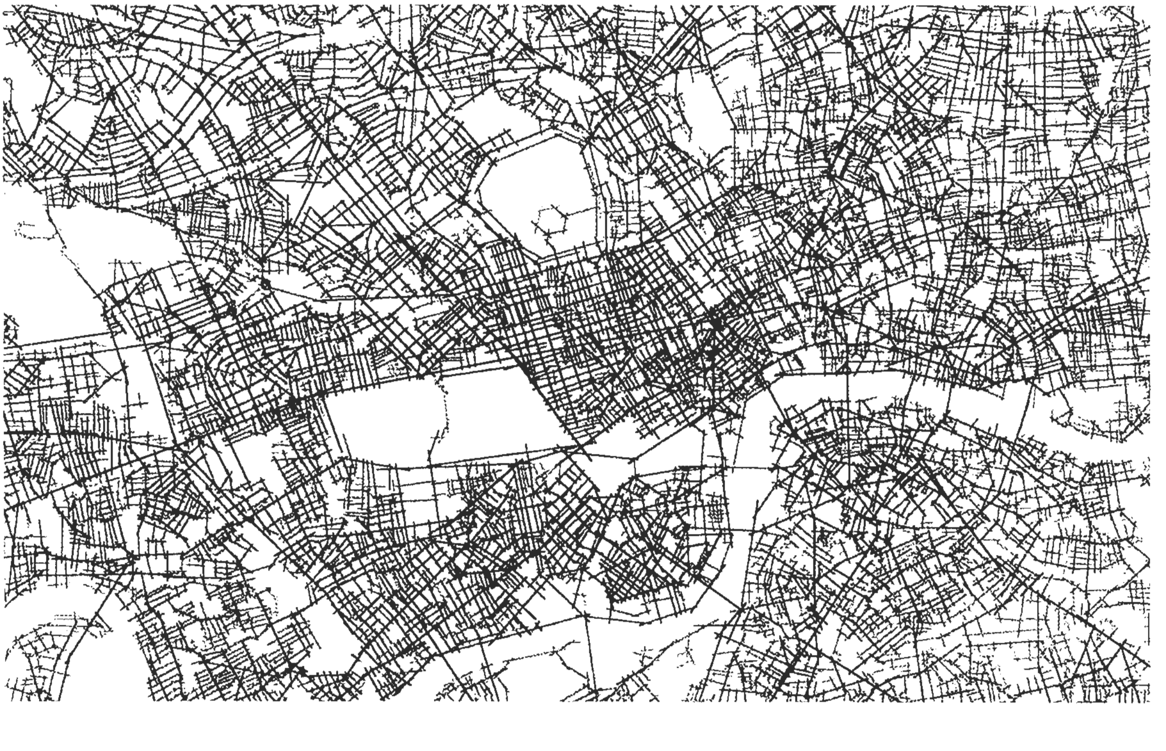
read the matrix of possible routes according to a larger-scale logic than pedestrians.[7]

The principle of natural movement

This relationship between the structure of the urban grid and movement densities along lines can be called the principle of 'natural movement'. Natural movement is the proportion of movement on each line that is determined by the structure of the urban grid itself rather than by the presence of specific attractors or magnets. This is not initially obvious, but on reflection does seem natural. In a large and well-developed urban grid people move in lines, but start and finish everywhere. We cannot easily conceive of an urban structure as complex as the city in terms of specific generators and attractors, or even origins and destinations but we do not need to because the city is a structure in which origins and destinations tend to be diffused everywhere, though with obvious biases toward higher density areas and major traffic interchanges. So movement tends to be broadly from everywhere to everywhere else. To the extent that this is the case in most cities, the structure of the grid itself accounts for much of the variation in movement densities.

(c)

(d)



Figs 4c & d.

(e)

Fig. 4e.

We should then expect that the distribution of line intensity in axial maps will foreshadow densities of moving people. Because the line intensities are really rough indices of precise numerical values, this proposition can of course be tested by selecting areas and correlating movement rates against integration values. However, because movement along a particular line is influenced in the main by its position in the larger-scale urban grid, we must take care to include enough of the whole urban grid in our analysis to ensure that each line in the area we are studying is embedded in all the urban structure that may influence its movement. We cannot then do better than to begin with the whole of an urban system, or at least a very large part of it in order to ensure that our study area is sufficiently well embedded.

In order to analyse an area in inner London, then, we begin with an axial representation of the very large part of London shown in Fig. 5, which covers the area approximately within the North and South Circular Roads. Figure 4c–e is then a series of analyses of integration at different radii. Figure 4c is the radius-n analysis, and as such shows the most global structure of London, with a strong edge-to-centre pattern centred on Oxford Street, which is the most integrated line. Figure 4d is the radius-3 analysis, which highlights a much more localized structure, including most local shopping streets, but also picks out Oxford Street as the dominant integrator. This implies that Oxford Street is not only the strongest global integrator in London as a whole, but also the strongest local integrator of its surrounding area. Figure 4e is a radius-10 (or radius-radius) analysis, meaning that the integration analysis is set at the mean depth of the whole system from the main integrator, which in this case is 10. The effect of setting the radius of analysis at that of the main integrator is that each line is analysed at the same radius which is at the same time the maximum radius possible without differences in radius between lines. The effect of

Fig. 5. Axial map of Greater London within the North and South Circular Roads.

a radius-radius analysis is to maximize the globality of the analysis without inducing 'edge effect', that is the tendency for the edges of spatial systems to be different from the interior area because they are close to the edge. Taken together, the figures show a remarkably true-to-life functional picture of London as a whole, highlighting all the main in and out routes and shopping high streets.

The reason that a spatial analysis can give such a true-to-life functional picture is due to the powerful influence that natural movement – the tendency of the structure of the grid itself to be the main influence on the pattern of movement – has on the evolution of the urban pattern and its distribution of land uses. To test this properly we must translate back from graphics to numbers. Figure 6a selects a small area within the system, more or less coterminous with the named area of Barnsbury, and assigns precise 'integration values' to each line. Figure 6b then indexes observed movement rates of adult pedestrians on each line segment throughout the working day.[8] Figure 6c is a scattergram plotting pedestrian movement rates against radius-3 integration. The R-squared value shows that about three-quarters of the differences between line segments in their movement rates are due to their configurational position in the larger-scale grid. Note, by the way, that we are still calculating integration with respect to a much larger system than that shown in Fig. 6a. Movement is not only largely determined by configuration, but also by configuration on a fairly large scale.

Readers can consult published texts for detailed results, but similar results have been achieved across a great range of studies, and even better (though slightly different) results have been

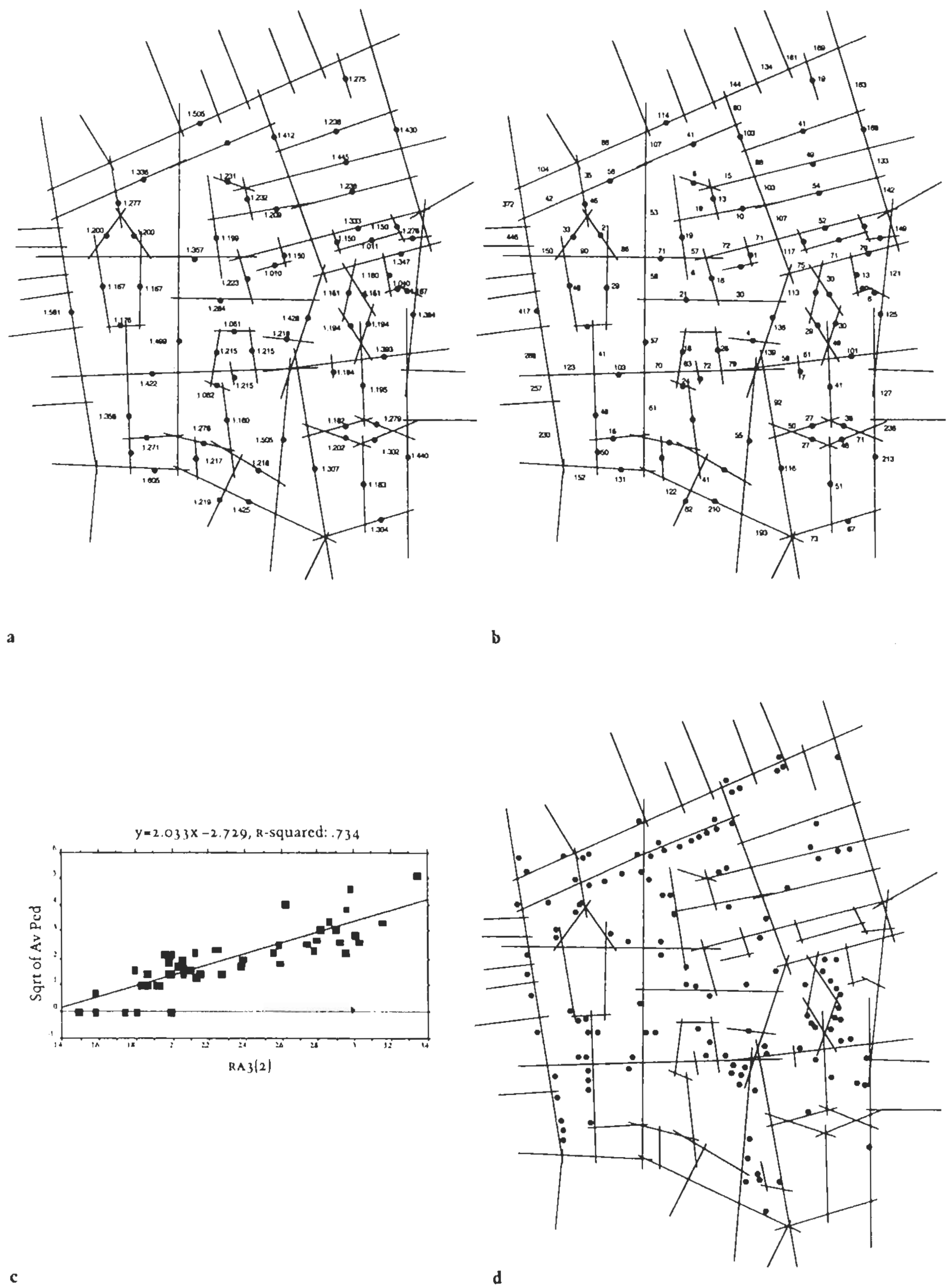


Fig. 6. a, b, c, d.

found from studies relating vehicular movement to spatial configuration.[9] These studies show that the distribution of pedestrian movement in the urban grid is to a considerable extent determined by spatial configuration, with the actual levels also strongly influenced by area building densities (though the effects of building density are not in general found at the level of the individual line), while vehicular movement is strongly influenced by spatial integration in association with net road width, that is the width of the road less the permitted car parking. In the case of vehicular movement the second variable, net road width, does influence movement on a line-by-line basis and plays a more significant part in the larger scale road network.[10]

We may investigate another key component of successful urbanism, the informal use of open spaces for stopping and taking pleasure, by using a similar technique. Figure 7 is a 'convex isovist' representation of the City of London's few, informal open spaces, which vary remarkably in their degree of informal use. Attempts to account for the pattern of well and poorly used spaces in the City in terms of commonly canvassed explanations have been singularly unsuccessful. For example, some spaces hemmed in by traffic are several times better used than adjacent spaces without traffic, exposed spaces often perform better than spaces with good enclosure, some of the most successful spaces are in the shadow of tall buildings, and so on. The only variable that correlates consistently with the degree of use of observed informal spaces is, in fact, a measure of the 'Roman property', noted in Fig. 2c, which we call the 'strategic value' of the isovist. This is calculated by summing the integration values of all the lines which pass through the body of the space (as opposed to skirting its edges). This makes intuitive sense. The primary activity of those who stop to sit in urban spaces seems to be to watch others pass by. For this, strategic spaces with areas close to, but not actually lying on, the main lines of movement are optimal. The main fault in most of the modern open spaces we have observed (with the most notable exception of Broadgate, which has the most successful spaces in the City of London) is that the designers have given too much attention to local enclosure of the space, and too little to strategic visual fields – yet another instance of an overly localized view of space. The general rule seems to be that a space must not be too enclosed for its size. The visibility field must be scaled up in proportion to the scale of the space.

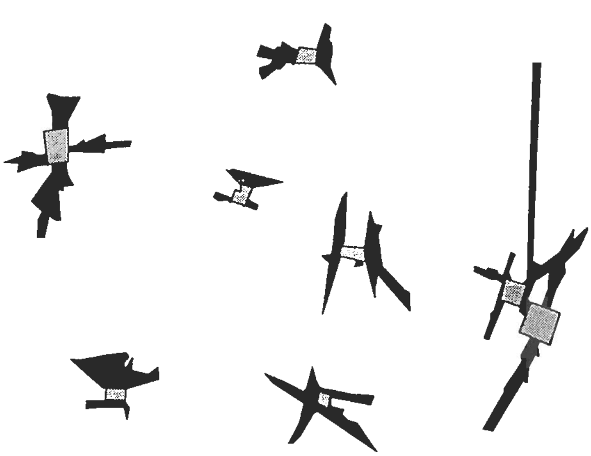


Fig. 7. Convex isovists from eight City of London squares.

Once we have the trick of correlating numbers indexing observed function with numbers indexing spatial patterns we can extend it to anything that can be represented as a number and located in space. When we do so, it turns out that everything seems to relate to space, and therefore to movement in some way: retail, building densities, indeed most types of land use seem to have some spatial logic which can be expressed as a statistical relation between spatial and function measures. Even crime can be spatially correlated. Figure 6d plots burglaries within a twelve-month period in the Barnsbury area. Visually, it looks as though there may be some effect from configuration, in that the densest concentrations seems to be in less integrated locations, while some of the more integrating lines are relatively free. Is this true? By assigning each dwelling the integration value of the line on which it opens we can ask if burgled dwellings are significantly more segregated or integrated than unburgled dwellings. It turns out that burgled dwellings are significantly more segregated on average than unburgled dwellings.

Now let us look at other aspects of how things are distributed in the urban grid. Take, for example, the well-known Booth map on London, in which socioeconomic classes are plotted from gold for the best off, through to red for merchant grade houses, then through pink to grey and black for the poorest. The most integrated streets are lined with red, and as you move into the less important, and less integrating streets, the grade of housing falls off, leaving the poorest in the most segregated areas. There is also a subtler organization concealed in the Booth map, one which provides an important clue to one of the hidden secrets of urban space: how different uses and economic classes are mixed in the same area by using a principle that can be summarized as 'marginal separation by linear integration'. If we look carefully we can see that different grades of housing – and in other situations we will find different land uses – may often be in close proximity but separated effectively by being on different alignments, often as part of the same urban block. The fundamental land use element is not the zone or even the urban block but the line: land use changes slowly as you progress along particular lines of movement, but can change quite sharply with ninety-degree turns onto different alignments. Since we know that the pattern of alignments is the fundamental determinant of movement, we can begin to see that the structure of the urban grid, the distribution of land uses, and built form densities are in the historically evolving city bound up with each other in a dynamic process centred on the relation of the grid structure to movement.

Which then is primary? Let us argue this through the spatial distribution of retail, the commonest nonresidential land use. We may already have been suspected of having confused the effects of spatial configuration on movement with the effect of shops. Are not the shops the main attractors of movement? And do they not lie on the main integrators? This is of course true. But it does not undermine what is being said about the structure of the grid as the prime determinant of movement. On the contrary it makes the argument far more powerful. Both the shops and the people are found on main integrators, but the question is: why are the shops there? The presence of shops can attract people but they cannot change the integration value of a line, since this is purely a spatial measure of the position of the line in the grid. It can only be that the shops were selectively located on integrating lines, and this must be because they are the lines which naturally carry the most movement. So, far from explaining away the relation between grid structure and movement by pointing to the shops, we have explained the location of the shops by pointing to the relation between grid and movement.[11]

Now of course in a sense to say this is to say the obvious. Every retailer knows that you should put the shop where people are going to be anyway, and it is no surprise if we find that the structure of the urban grid influences at least some land uses as it evolves. It would be surprising if it were not the case. However, a little more than this is being claimed. It is being suggested that there is an underlying principle which, other things being equal, relates grid structure to movement pattern not only on the main lines in and out of a city, but also in the fine structure, and through this gives rise to a whole multiplicity of interrelationships between grid structure, land uses, densities, and even the sense of urban wellbeing and fear.

Multiplier effects and the movement economy

We can pursue this by thinking carefully about what it would take to produce this degree of agreement between grid structure, movement, land uses and densities. We find ourselves unavoidably led towards a theory of the general formation of the city through the functional shaping of its space by movement. Let us begin by considering that. An urban system, by definition, is one which has at least some origins and destinations more or less everywhere. Every trip in an urban system has three elements: an origin, a destination, and the series of spaces that are passed through on the way from one to the other. We can think of passage through these spaces as the by-product of going from *a* to *b*. We already know that this by-product, when taken at the aggregate level, is determined by the structure of the grid, even if the location of all the *a*'s and *b*'s is not.

Location in the grid therefore has a crucial effect. It either increases or diminishes the degree to which the movement by-product is available as potential contact. This applies not only to individual lines, but to the groups of lines that make up local areas. Thus there will be more integrating and less integrating areas, depending on how the internal structure of the area is married into the larger-scale structure of the grid, and this will mean also areas with more by-product and areas with less.

Now if cities are, as they were always said to be, 'mechanisms for generating contact', then this means that some locations have more potential than others because they have more by-product and this will depend on the structure of the grid and how they relate to it. Such locations will therefore tend to have higher densities of development to take advantage of this, and higher densities will in turn have a multiplier effect. This will in turn attract new buildings and uses, to take advantage of the multiplier effect. It is this positive feedback loop built on the relation between the grid structure and movement which gives rise to the urban buzz, which we prefer to be romantic or mystical about, but which arises from the coincidence in certain locations of large numbers of different activities involving people going about their business in different ways. Such situations invariably arise through multiplier effects generated from the basic relation between space structure and movement, and ultimately this depends on the structure of the urban grid itself. In other words, how the urban system is put together spatially is the source of everything else.

We may illustrate this negatively through a notorious case where the urban buzz does not occur, in spite of the coexistence in a small area of many major functions. The example is the area of the South Bank cultural centre in London, where within a few hundred metres can be found Europe's largest and most diverse cultural complex, a major international railway terminus, extensive office development, significant residential development and a famous riverside walk. Why do all these facilities not add up into an urban area with the qualities called for by these high-level facilities? It can only be the way it is put together. This is indeed the case. Our studies have shown that each of the various constituencies of space users – travellers, residents, office workers, tourists, concert goers and gallery visitors all use space in a different way and, as it were, move through the area largely on separate routes passing each other like ships in the night. It is the failure of the configuration of space to bring these different constituencies into patterns of movement and space use where all are prioritizing the same space, that deprive the area of the multiplier effects that occur when different constituencies of space use all spark off each other.

If these arguments are right, it means that all the primary elements of urban form – the structure of the urban grid, the distribution of land uses, and the assignment of development densities – are bound together in the historical city by the principle that relates the structure of the urban grid to the by-product of movement. It means that under certain conditions of density and integration of the grid structure things can happen that will not happen elsewhere. Movement is so central to this process that we should forthwith cease to see cities as being made up of fixed elements and movement elements and instead see the physical and spatial structure as being bound up to create what we have called the 'movement economy', in which the usefulness of the by-product of movement is everywhere maximized by integration in order to maximize the multiplier effects which are the root source of the life of cities.

Urbanity, we suggest, is not so mysterious. Good space is used space. Most urban space use is movement. Most movement is through movement, that is, the by-product of how the grid offers routes from everywhere to everywhere else. Most informal space use is also movement related, as is the sense and fact of urban safety. Land uses and building density follow movement in the grid, both adapting to and multiplying its effects. The urban buzz, or the lack of it when it suits us, is the combination of these, and the fundamental determinant is the structure of the grid itself. The urban grid through its influence on the movement economy is the fundamental source of the multifunctionality that gives life to cities.

Parts and wholes

We can also show how the movement economy creates the part–whole structure of cities. We have already noted that movement occurs at different scales: some localized and some more globalized. Long journeys will tend naturally to prioritize spaces which are globally more integrated, more local journeys those which are more locally integrated. The space system is literally read – and readable – at a different scale. Since different radii of integration reflect different scales of the urban system, it will turn out that the key to understanding parts and whole is understanding the relations between the different radii of integration.

Consider, for example, the relation between the City of London and London as a whole. Figure 8a is a close-up of the axial map of the City of London in context. Figure 8b is a scattergram plotting each line in the London axial map as a whole as a point located according to its degree of global (radius-n) integration on the horizontal axis and its degree of local (radius-3) integration on the vertical axis. The dark points are the lines which make up the City of London. The dark points form a good linear scatter about their own (invisible) regression line, and cross the main regression line at a steeper angle. The linearity implies a good relation between local and global integration, the steeper slope across the regression line implies that the most integrated lines within the city, which are the lines from the outside towards the centre, are more locally than globally integrated. Their local integration is as it

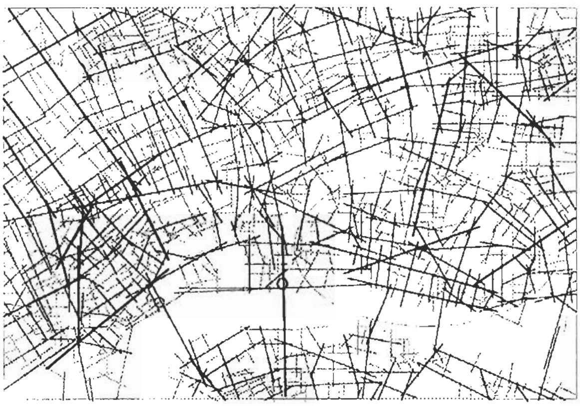


Fig. 8a. City of London within the context of Greater London.

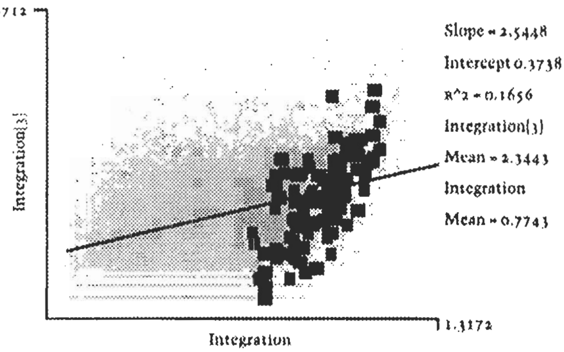


Fig. 8b. Scatter of the City of London within the context of Greater London.

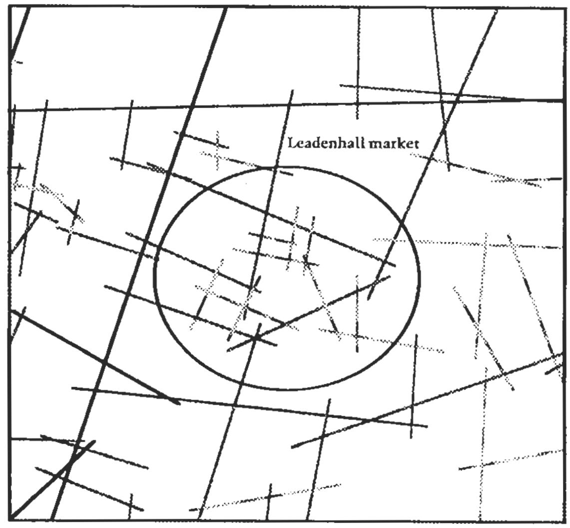


Fig. 8c. Leadenhall market, City of London.

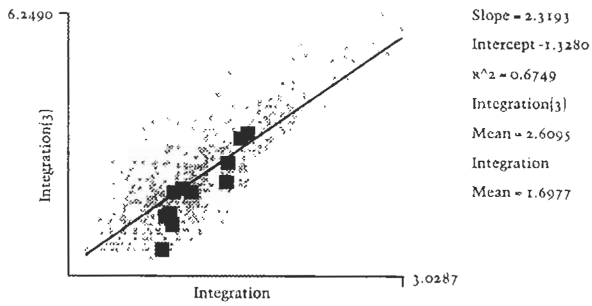


Fig. 8d. Scatter (in black dots) of Leadenhall market within the context of the City of London.

were intensified for their degree of global integration. Repeating this experiment with all of the well-known named London areas, such as Soho, Covent Garden, Bloomsbury and even Barnsbury, yields this kind of scatter. In other words, the relation of part and whole in the axial map is made up at least in part of the relation between local and global integration. The reason this is so is that each local area has its heart linked to the supergrid lines that surround it by strong integrators. These form an edge-to-centre structure in all directions, and the less-integrated areas are within the interstices formed by the structure. The strong local integrators which define the slope of the dark points for the local area are invariably these edge-to-centre lines.[12]

Remarkably, we find exactly the same phenomenon on a much smaller scale, for example within the City of London. Figure 8c homes in on the Leadenhall Market area, and Fig. 8d shows the City scatter with the Leadenhall Market area as the dark points. Once again we find the local area effect. The effect of this is that as you move down the supergrid lines – Gracechurch Street or Leadenhall Street – then Leadenhall Market is available as a well-structured local intensification of the grid, itself laid out in much the same way as the town is laid out. Once you are near it in the adjacent streets, it becomes a powerful attractor.

We can draw a simple conclusion from these results, one which I believe agrees with intuition: that the more the set of dark points forms a line crossing the regression lines for the whole city but tending to greater steepness, there is more local integration than global, then the more the subarea is distinctive; while the more the dark points lie on the City regression lines, the more they are simply sets of smaller spaces related to the main grid, but not forming a distinctive subarea away from it. This depends, however, on the dark points themselves forming a good line, since without that we do not have a good integration interface – that is, a good relation between the different scales of movement – in the first place, regardless of where it is in relation to the main City. It depends also on the dark points including points well up the scale of integration. A clutch of points (bottom left) will be very segregated, and not function as a subarea.

We have found an objective spatial meaning, it seems, to the areas we name as areas, and in such a way as to have a good idea of the functional generators of their distinctive urban patterns. We have a key to how at least some cities can be put together as cities of parts without losing the sense of the whole. Historically, it seems, cities exploited movement constructively to create dense, but variable, encounter zones to become what made them useful: 'mechanisms for generating contact'. They did this by using space to generate multiplier effects on the relation between movement and encounter. This was achieved by quite precise spatial techniques, applied now this way now that (for example, in Arabic cities we find a quite different development of the same underlying laws), but always having the effect of creating well-defined relationships between different levels of movement: between the movement within buildings and the movement on the street, between localized movement in less important streets and more globalized pattern of movement, and between the movement of inhabitants and the movement of strangers entering and leaving the city. In a sense, cities were constructed to be, in the words of Dr John Peponis,[13] interfaces between scales of movement.

The interface between different radii of integration was the spatial means to the functional end. It created a close relation between more localized and more globalized movement. It is, therefore, the key to the local by-product effect, and the means to create local advantage from global movement. The spatial technique to achieve this was to maintain a number of spatial interfaces: between building entrances and all spaces, at whatever scale; between smaller spaces and the larger urban scale through the relation between the convex and linear structures; and between different scales of the linear structure, especially between parts and the whole.

Disurbanism

The urban movement economy, arising from the multiplier effect of space, depends on certain conditions: a certain size, a certain density, a certain distribution of land uses, a specific type of grid that maintains the interface between local and global, and so on. Once this is spelled out, it is easy to see how thoroughly some of our recent efforts have disrupted it, so much so that we must think of many developments of recent years as an exercise in the spatial technique of disurbanism. 'Disurbanism' is intended to convey the reverse of the urban spatial techniques we have identified; the breaking of the relation between buildings and public space; the breaking of the relation between scales of movement; and the breaking of the interface between inhabitant and stranger.

Consider, for example, the integration map of an area around Barnsbury, which includes three housing estates around the Kings Cross railway lands site (the empty area), as in Fig. 9a. The estates are easy to pick out: they are more complex and at a smaller spatial scale than the surrounding street-based areas, and each is marked by its density of light shaded, that is segregated, lines. If we try to plot these estates as dark point scatters of local against global integration, as in Fig. 9b, c and d, then we find that in each case the estate scatter forms a series of layers, each distributed in a more or less vertical pattern. Here we note three consequences of this type of spatial design. First, the estate is substantially more segregated than the rest of the urban surface and, what is more problematic, segregated as a lump. Good urban space has segregated lines, but they are close to integrated lines, so that there is a good mix of integrated and segregated lines locally. Second, there is a poor relation between local and global integration, that means a very unclear relation between the local and global structure. Third, the scatter does not cross the line to create a well-structured local intensification of the grid.

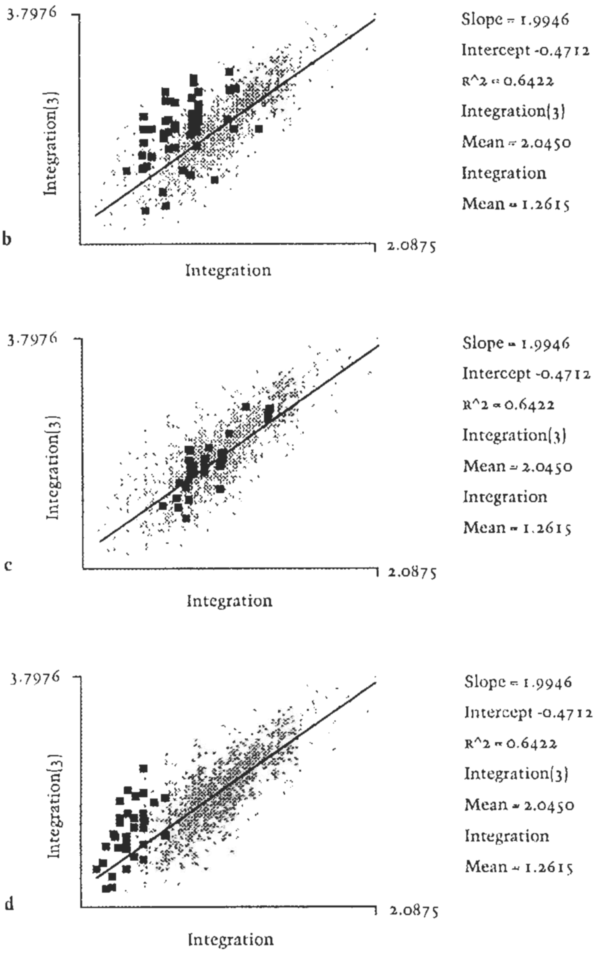
What this means in functional terms is that all interfaces are broken: between building and public space; between localized and less localized movement; and between inhabitant and stranger. Of course life is possible in such a place. But there is now evidence to suggest that we ought to be more pessimistic. Efforts to trace

a

Fig. 9a–d. Global integration map of King's Cross with three housing estates picked out in black and the scatters (in black dots) of the three housing estates within their larger context.

the effects that such designs can have over a long period on the type of life that goes on in them suggest that there is a pattern of long term development in which spatial designs create serious lacunas in natural movement, which then attract anti-social uses and behaviours. In extreme cases, where the lacunas of natural movement are the integration core of the estate itself, then the situation may become pathological.

These 'disurban' places arise from a poorly structured local configuration of space; as a consequence of which the main elements of the movement economy are lost. A similar pattern of loss can also arise through dispersion. If we move from an urban system that is dense and nucleated to one that is dispersed and fragmentary, it is obvious that the mean length of journeys will, other things being equal, increase. It is less obvious, but equally true, that the by-product effect will also be diminished. As dispersion increases, it becomes less and less likely that connected locations will benefit from the by-product of movement. In effect, as dispersion increases, the movement system



Figs 9b–d.

becomes more like a pure origin–destination system. Instead of one journey accomplishing a number of purposes, more journeys, each one accomplishing fewer purposes, must be made to attain the same goals. These are the basic reasons why people travel farther in the country, and why most of this extra travel is in private cars.[14]

A similar effect can arise even in a comparativly dense urban system through an urban design policy of replacing continuous urban structure with specialized enclaves. This will also tend to eliminate by-product. Enclaves are, almost by definition, destinations which are not available for natural movement. They form discontinuities in the urban grid. Because this is so they are in many ways comparable in their effects to physical dispersion, and similarly disruptive of the movement economy. Any tendency in an urban structure towards 'precinctization' must also be a tendency towards a lessening of the useful by-product, and therefore of the multiplier effect on which urban vibrancy depends.

These arguments suggest that the culturally sanctioned values that are embedded in attitudes towards urban design that until quite recently were taken for granted – lowering densities wherever possible, breaking up urban continuity into well-defined and specialized enclaves, reducing spatial scale, separating and restricting different forms of movement, even restricting the ability to stop travellers from moving and taking advantage of the by-product effect – are fundamentally inimical to the natural functioning of the city and its movement economy. It is not density that undermines the sense of well-being and safety in urban spaces, but sparseness; not large spatial scale, but its insensitive reduction; not lack of order but its superficial imposition; not the 'unplanned chaos' of the deformed grid, but its planned fragmentation. Without an understanding of the spatial and functional nature of the city as a whole, we are in danger of eliminating all the properties of density, good spatial scale, controlled juxtaposition of uses, continuity, and integration of the urban grid on which the well-ordering and well-functioning city depends.

Reflections on the origins of urbanism and the transformation of the city

These conclusions can only reinforce the thought with which we began: our interventions in the city can only be based on our understanding of the city. Where this understanding is deficient, the effects can be destructive, and this will be more the case according to the degree that this false understanding is held in place by a value system. The value system according to which we have been transforming our cities over much of the past century has always appeared as a kind of urban rationality, but it was never based on the study of the city. Where then did it come from?

Let us first reflect a little on the nature and origins of cities, why we have them and what made them possible. Towns, as physical objects, are clearly specialized forms of spatial engineering which permit large number of people to live in dense concentrations without getting on each other's nerves, and minimize the effort and energy needed for face-to-face contact with each other and with the providers for needs. Towns, we suggest, were in fact made functionally possible in the first instance by a transmutation in the way energy flowed through society. It is most easily explained through the geographer Richard Wagner's distinction between two kinds of energy-related artifact: *implements* which transmit or accelerate kinetic energy, and *facilities* which store up potential energy and slow down its transfer.[15] For example, a flint knife is an implement, whereas a dam is a facility. Whatever else made towns possible, there is no doubt that they were usually marked by a radical increase in facilities, most especially irrigation systems and food storage facilities.

What made towns possible socially was an invention we are so familiar with that we tend to take it for granted and forget it is there: the urban grid. The urban grid is the organization of groups of contiguous buildings in outward-facing, fairly regular clumps, amongst which is defined a continuous system of space in the form of intersecting rings, with a greater or lesser degree of overall regularity. Urban grids were never inevitable. In fact, the archaeological record reveals many proto-towns with quite different morphologies.

The urban grid was, however, the first powerful theorem of urban spatial engineering. Its crucial characteristic is that it is itself a facility – one that takes the potential movement of the system and makes it as efficient and useful as possible. The grid is the means by which the town becomes a 'mechanism for generating contact', and it does this by ensuring that origin–destination trips take one past outward-facing building blocks *en route*. That is, they allow the by-product effect to maximize contact over and above that for which trips are originally intended.

In the nineteenth century, however, under the impact of industrialization and rapid urban expansion, two things happened. First, to cope with sheer scale, the urban spatial grid was thought of as more of an implement than a facility. That is, it was seen as a means to accelerate movement in order to overcome size. Alongside this it was envisaged as a set of point-to-point origins and destinations, rather than as an 'all points to all points' grid, which is the product of an urban movement economy. Second, the city began to be seen not as a grid-based civilization, but as the overheated epicentre of focal movement into and out of the city, and as such the most undesirable of locations. A social problem was seen in the disorderly accumulation, in and around city centres, of people brought in to serve the new forms of production. Big became synonymous with bad, and density became synonymous with moral depravity and political disorder. It was this that gave rise to much of the value system of nineteenth-century urban planning, as well as the more extreme proposals for the dispersion and ruralization of the city and its population.

Unfortunately, much of this nineteenth-century value system survived into the twentieth century, not so much in the form of consciously expressed beliefs and policy objectives as in assumptions as to what constituted the good city. For much of the twentieth century, nineteenth-century anti-urbanism provided the paradigm for urban design and planning. It would be good to believe that this may have now changed, and that cities are again being taken seriously. But this is not the nature of human beliefs when they become embedded in institutional forms and structures. Many aspects of the nineteenth-century urban paradigm have not yet been dismantled, and are still to be found enshrined in everyday policies towards density, in novel ways of breaking up urban continuity into well-defined and specialized enclaves, in continuing to reduce spatial scale, and in separating and restricting different forms of movement. These relics of an outdated paradigm do not derive from an understanding of cities. On the contrary, they threaten the natural functioning and sustainability of the city.

Notes

1 The best recent review of these issues is Owens, S. (1992) Land-use planning for energy efficiency, *Applied Energy*, **43**, 1–3, Special issue on the rational use of energy in urban regeneration, R. Hackett and J. Bindon (eds), Elsevier Applied Science. An important source on settlement forms on which she draws is P. Rickaby (1987) Six settlement patterns compared, *Environment & Planning B: Planning & Design*, **14**, 193–223. Significant recent contributions include Banister, D. (1992) Energy use, transport and settlement patterns, in M. Breheny (ed.) *Sustainable Development and Urban Form*, Pion; also Hall, P. (1994) Squaring the circle; can we resolve the Clarkian paradox? *Environment & Planning B: Planning & Design*, **21**, s79–s94.

2 For a discussion see Batty, M. (1989) Urban modelling and planning: reflections, retrodictions and prescriptions, in B. Macmillan (ed.) *Remodelling Geography*, Basil Blackwell, Oxford, pp. 147–169. See also Batty, M. and Longley, P. (1994) *Fractal Cities*, Academic Press, London.

3 Hillier, B. *et al.* (1993) Natural movement: or configuration and attraction in urban pedestrian movement, *Environment & Planning B: Planning & Design*, **20**; and Penn, A. and Dalton, N. (1994) The architecture of society: stochastic simulation of urban movement, in N. Gilbert and J. Doran (eds) *Simulating Societies: The Computer Simulation of Social Phenomena*, UCL Press, London, pp. 85–125.

4 In this sense, it is an instance of what Ian Hacking calls 'the creation of phenomena', which then leads to the evolution of theory: Hacking, I. (1983) *Representing and Intervening*, Cambridge University Press, pp. 220–32.

5 The figures are taken from a case study carried out by Marios Pelekanos while a student on the MSc in Advanced Architectural Studies at the Bartlett School of Graduate Studies, UCL, in 1989.

6 Hillier, B. *et al.* 'Natural movement'. Op. cit.

7 Penn, A. *et al.* (1995) Configurational modelling of urban movement networks (submitted for publication, currently available from the Bartlett School of Graduate Studies).

8 In this study, each line segment was observed in total for about 50 minutes, spread during five different time periods: 8–10 am, 10–12 noon, 12–2 pm, 2–4 pm and 4–6 pm. The data is therefore of very high quality. Experiments have shown however, that comparatively short periods of observation can be sufficient where there are reasonable numbers of people to be observed. In sparse environments, more protracted observations are required.

9 See for example Penn, A. and Hillier, B. *Configurational modelling* (see note 7).

10 Penn, A. and Hillier, B. (see note 7).

11 This issue is discussed in greater detail in Hillier, B. *et al.* (1993) Natural movement (see note 3).

12 This structure has also been found in small towns and called a 'deformed wheel', since there is always a semigrid, or hub, of lines near the centre, strong integrators which link this semi-grid to the edges, like spokes, and some edge lines are also integrated, forming a partial rim. This structure is usually the main public space structure, while less integrated residential areas form in the interstices form by the wheel. See Hillier, B. (1989) The architecture of the urban object, *Ekistics*, Special issue on space syntax: social implications of urban layouts, 5–21.

13 Dr John Peponis of the Georgia Institute of Technology and the Polytechnic University of Athens, in conversation.

14 See, for example, Department of Transport, National Transport Survey (1983) 1978/79 Report, HMSO, Norwich, Table 10.4, p. 71. (See also NTS: 1975/76 Report, Table 3.17, p. 37.)

15 Wagner, P.L. (1960) *The Human Use of the Earth*, New York, Chapter 6. For a further discussion see Flannery, K. (1972). The origins of the village as a settlement type in Mesoamerica and the Near East: a comparative study, in P. Ucko *et al.* (eds) *Man. Settlement and Urbanism*, Duckworth, pp. 23–53.