Part three
The laws of the field

3

Chapter eight

Is architecture an ars combinatoria?

Despite the merits of rectangular dissections as models of smaller plans, there is an increasing proportion of 'theoretical possibilities' for larger dissections which nevertheless become rather unlike the plans of buildings, and hence begin to lose their practical interest. Such dissections consist, certainly, of rectangular components corresponding to rooms, packed together in different configurations. But these configurations are not at all probable architecturally, in ways which are hard to pinpoint precisely, but are no less real for that. It is something to do with such facts as that real buildings tend to have limited depth, because of the needs of daylighting and natural ventilation, so that when large they become organised into regular patterns of wings and courts. Or that rooms are set along relatively simple and coherent circulations systems consisting of a few branching corridors which extend along the buildings' whole length. There are many dissections which are made up, by contrast, of a deep maze like agglomeration of overlapping rectangles, many of them completely internal and through which any linking pattern of circulation routes would be circuitous and confusing. If we could capture properties like these in explicit geometrical measures, then we might be able to limit the study of dissections, for example, to a much reduced class of arrangements which would all be 'building-like' in some well defined sense. **Steadman, 1983**

The deepest root of the trouble lies elsewhere: a field of possibilities open into infinity has been mistaken for a closed realm of things existing in themselves
Herman Weyl

Endless corridors and infinite courts

No idea in the theory of architecture is more seductive than that architecture is an *ars combinatoria* – a combinatorial art: the idea that the whole field of architectural possibility might be made transparent by identifying a set of basic elements and a set of rules for combining them so that the application of one to the other would generate the architectural forms which exist, and open up possibilities that might exist and be consistent with those that do. By showing architectural forms to be a system of transformations in this way, the elements and rules would be held to be a theory of architectural form – the system of invariants that underlie the variety to be found in the real world. The best-known statement of this hope is that of William Lethaby when he calls for 'a true science of architecture, a sort of architectural biology which shall investigate the unit cell and all possibilities of combination'.[1]

At first sight, this seems promising. Most buildings seem to be made up from a rather small list of spatial elements such as rooms, courts and corridors, which vary in size and shape but which are usually found in fairly familiar arrangements: corridors have rooms off them, courts have rooms around them, rooms may connect only with these or may also connect directly to each other to form sequences, and so on. Similarly, the aggregates of buildings we call villages, towns and cities seem to be constructed from a similarly small and geometrically well-defined lexicon of streets, alleys, squares, and so on. With such an encouraging start, we might hope with a little mental effort to arrive at an enumeration of the combinatoric possibilities in the form of a list of elements and the possible relationships they can enter into so that we can build a reasoned picture of the passage from the simplest and smallest cases to the largest and most complex.

Unfortunately, such optimism rarely survives the examination of real cases. If, for example, we consider the cross-national and cross-temporal sample of 177 building plans brought together in Martin Hellick's *'Varieties of Human Habitation'*,[2] we may well feel inclined to confirm at a very broad level – and with great geometric variation – the idea that there are certain recurrent spatial types such as rooms, courts and corridors, but we also note the prodigious variations of overall layout which seem to be consistent with each. The historical record of actual buildings and how they have evolved suggests that most buildings are morphologically unique, and it is far from obvious how any combinatorial approach could reduce them to a list of types.

Even if we isolate the problem of spatial relations from that of shape and size by, for example, analysing plans as graphs, then we still find cornucopian variety rather than simple typology. For example, a recent study of over 500 English vernacular houses built between 1843 and 1930 reveals exactly six pairs of duplicate graphs, even though the sample was taken from a single country during a period where some typological continuity could be expected.[3] Plans seem to be individual, often with family resemblances or common local configurations, but rarely consistent enough or clear enough to suggest a simple division into types.

Theoretical investigations of architectural possibility have led to an even greater pessimism. For example, studies which have attempted to enumerate architectural possibility, even within artificially constrained systems such as the dissection of rectangles into patterns of room adjacencies,[4] have invariably shown that at an early stage in the enumeration the number of possibilities quickly outstrips the number of conceivable cases, and a combinatorial explosion of such violence is encountered as to exclude any practical possibility of continuing from smaller to larger systems. Thus Steadman concludes in his review of modern attempts at the systematic enumeration of building plans that '...for values of n (the number of cells in a rectangular "dissection") much greater than 10, the extent of combinatorial variety becomes so great that a complete enumeration is of little practical purpose; and indeed that for values of n not much larger than this, enumeration itself becomes a practical impossibility'.[5]

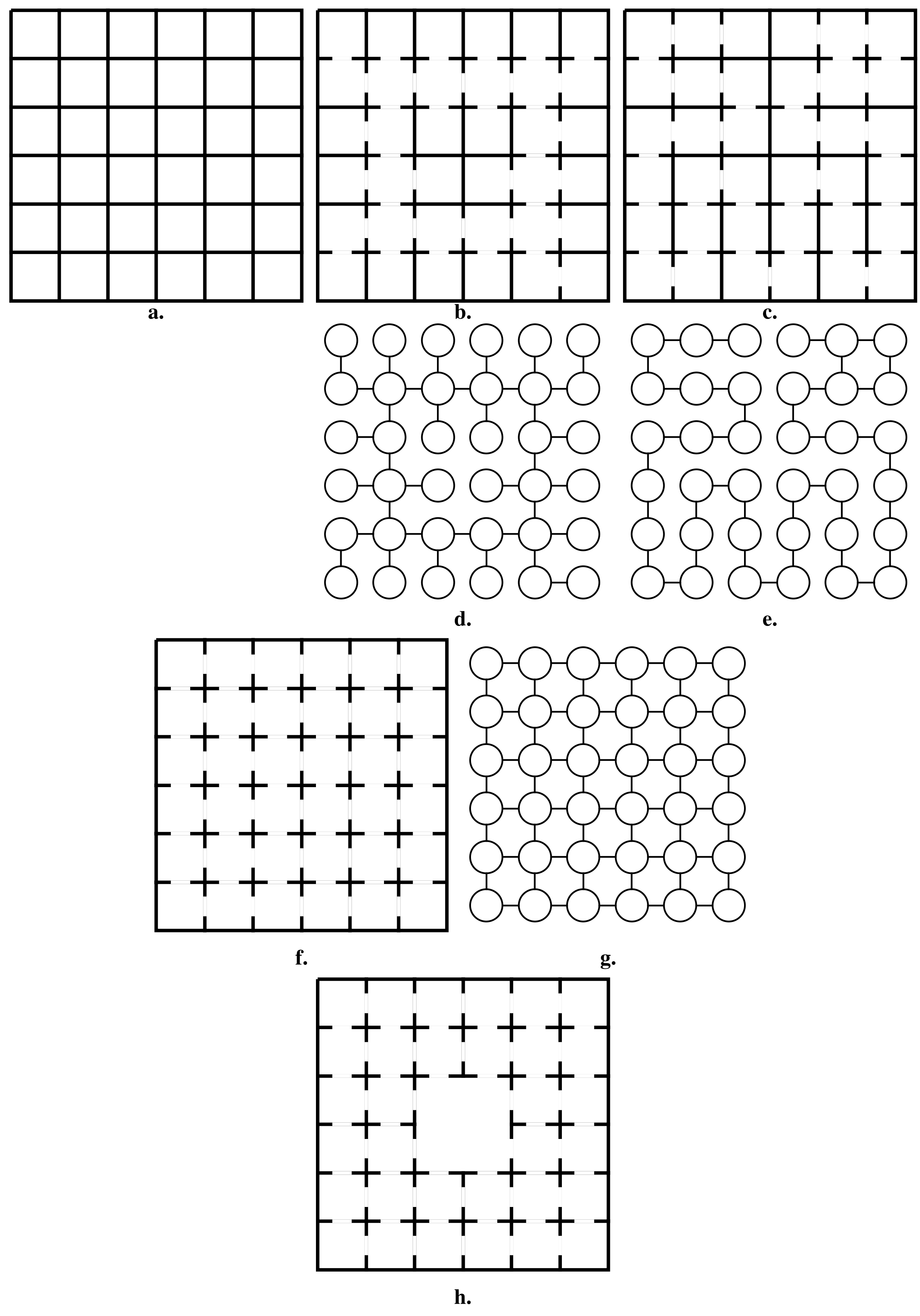
There are in fact strong *a priori* grounds for Steadman's caution. Although by circumscribing what we mean by a building in unlifelike ways, for example, by dealing only with rectangular envelopes, or by standardising the size and shape of spaces, one can place limits on combinatorial possibility to the point where we can in principle count numbers of possible arrangements, however large, the more constraints one places on the combinatoric system, the less we seem able to account for the variety which actually exists. But if we relax these constraints, it is far from obvious that there are any numerical limits at all on architectural possibility. For example, if we require all cells to be the same size then no cell can be adjacent to more than six others. But if we allow cells to vary in size and shape as much as necessary, then we may construct a corridor so that arbitrarily many cells are directly adjacent to it, or a court so that arbitrarily many cells are around it. Endless corridors and infinite courts must surely lead us to abandon simple cellular enumeration as a route to a combinatoric theory of spatial possibility in architecture.

P-complexes in a-complexes

There is in any case a further profound problem in the understanding of buildings as cellular dissections or aggregations. An arrangement of adjacent cells, whether arrived at by aggregation or subdivision, is not a building until a pattern of permeability from one cell to the other is created within it. For example, figure 8.1a shows a single adjacency complex, which we may call an a-complex, in which figures 8.1b and 8.1c inscribe different permeability complexes, or p-complexes. For clarity, the p-complexes of b and c are also shown as graphs in 8.1d and e.

Evidently, the two will be spatially very different buildings, even though the a-complexes are identical and each p-complex has exactly the same number of open and closed partitions. Over and above the question then, of how many a-complexes there are, we must therefore also ask how many p-complexes are possible within a given a-complex. We then find a second combinatorial explosion within the first: of possible p-complexes within a given a-complex. Although an a-complex whose graph is a tree (see Chapter 1) can only have one single p-structure inscribed within it (and then only if we disregard connections to the outside) as soon as this constraint is relaxed we begin to find the second combinatorial explosion: that of

Figure 8.1



the possible p-complexes within each a-complex.

Suppose, for example, that we start with a version of the 6×6 a-complex shown in figure 8.1a, in which each cell is demarcated from its neighbour by a two-thirds partition with a central doorway, as in figure 8.1f and g. Obviously, every time we close – or subsequently open – a doorway we will change the spatial pattern of the p-complex. The question is, how many ways are there of inscribing different p-complexes in this a-complex by closing and opening doors? We may work it out by simple combinatorial procedure. First we note that a regular n × m adjacency complex will always have (m(n–1)+(n(m–1)) internal partitions between cells, giving (6(6–1)+(6(6–1)) = 60 in this case. This means that the first time we select a door to close we will be making a choice out of 60 possibilities. The second will be out of 59, so there are 609, or 3540 possibilities for the first two doors. However half of these will be duplicates, since they differ only in the order in which the doorways were opened, so we need to divide our total by the number of ways there are of sequencing two events i.e. 60×59/1×2, or 1770. The third doorway will be chosen out of 58 remaining possibilities, so there will be 60×59×58 or 205320 possible combinations of three, but the number of duplicates of each will also increase to the number of different ways there are of ordering three events, that is 1×2×3 (= 6), so the total of different combinations for three doorways is 60×59×58/1×2×3 or 34220.

The total number of combinations for n doorways, will then be 60×59×58...×(60–n)/1 × 2× 3×...×n, or in general, n(n–1)(n–2)...(n–m)/m! In other words the number of duplicates increases factorially rising from 1, while the number of total possibilities is multiplied by one less each time. This means that as soon as m reaches n/2, then the number begins to diminish by exactly the same number that it previously expanded. The numbers in effect pass each other half way, so that there are the maximum number of different ways of arranging 30 partitions in 60 possible locations, but this number diminishes to 1 by the time we are opening the 60th doorway, just as it was when we opened the first doorway. These calculations reflect a simple intuitive fact, that once we have placed half the partitions, then what we are really choosing from then on is which to leave open, a smaller number than the partitions we have so far placed. When we have placed 59 partitions, there is only one location in which we can place the 60th, and this is why if we carry out the calculation at this point it will give a value of 1.

What exactly are the numbers we are talking about? The procedure we have outlined can in fact be expressed more simply in a well-known combinatorial formula which can be applied in any situation where we are assigning a given number of entities to a given number of possible assignments. If the number of doorways is d, and the number of partitions p, then the formula p!/d!(p–d)! will give us the number of possibilities which we have just worked out. With p=60, the highest value that the formula can yield will be when d is half the possible number, that is 60/2×30, and the result of the calculation 60!/(30!(60 – 30)!) is 118,264,581,600,000,000 (a hundred and eighteen thousand trillion). The second highest value, 114,449,595, 100,000,000, will be when d is 29 or 31, the next, 103,719, 935,500,000,000, when d is 28 or 32, and so on, and the lowest, 1, when d is 0 or 60, and the second lowest, 60, when d is 1 or 59. These kinds of numbers of possibilities, though quite modest by combinatoric standards, are almost impossible to grasp. To give an intuitive idea of the scale of possibilities we are dealing with in the modest complex, we might perhaps compare our maximum number of possible p-graphs for this comparatively small a-graph with another 18-digit number: the number of seconds believed to have passed since the big bang (provided it occurred 15 billion years ago), that is about 441,504,000,000,000,000. This means that if a computer had begun at the moment of the big bang to draw up all these possible configurations of doorways for this one modest adjacency complex, then it would have had to work at an average of one every four seconds to be finishing now. If we printed out the results on A4 sheets, and set them side by side, they would reach from Earth to the nearest star and back, or 141,255 times to the sun and back, or just short of a billion times round the world.

There are a number of ways of reducing these vast numbers. For example, each p-complex will have as many duplicates as there are symmetries in the system. We can therefore reduce all our totals by this factor. We may also decide that we are only interested in those p-complexes which form a single building, that is a complex in which each cell is accessible from all others without going outside the building. The maximum number of doors that can be closed without necessarily splitting the complex into two or more sub-complexes will always be (n-1)(m-1), or 25 in this case. No way is known of calculating how many of the p-complexes with 25 or less partitions will be single buildings, but, in any case, the realism of this restriction is doubtful because we have not so far taken any account of permeability to the exterior of the form, and in any case, a complex split into two is still a building complex and may be found in reality.

More substantively, we might explore the effects on imposing Steadman's 'light and air' restrictions on the form. Here we find they are far less powerful than we might think in restricting p-complexes. For example, we may approximate a form in which each cell has direct access to light and air by making an internal courtyard as in figure 8.1h give or take a little shifting of partitions to allow the inner corner cells direct access to the courtyard. Combinatorially, this has the effect of reducing the number of internal partitions by 4 to 56, and the maximum number that may be closed without splitting the building by 1 to 24. The number of p-complexes that can be inscribed within the a-complex is therefore still in the thousands of trillions.

We will find this is generally the case. The imposition of the requirement that each cell should have direct access to outside light and air makes relatively little impact on the number of p-complexes that are possible, the more so since direct access to external light and air will also mean an extra possible permeability in the system which we have not so far taken account of. It is clear that although light and air are inevitably powerful factors in influencing the a-complex, they place relatively little restriction on the possible p-complexes. We might even venture a generalisation. 'Bodily' factors like light and air have their effect on buildings by influencing the a-complex, but do not affect the p-complex which is determined, as we have seen in previous chapters, and as we will see more generally below, largely by the psycho-social factors which govern spatial configuration.

If we see buildings, as we must, as both physical and spatial forms, that is as a-complexes with p-complexes inscribed within them, then we must conclude that buildings as a combinatorial system take the form of one combinatorial explosion within another with neither being usefully countable except under the imposition of highly artificial constraints. Is the combinatoric question about architecture then misconceived? If it is, how then should we account for the fact that there do seem to be rather few basic ways of ordering space in buildings. What we must do, I suggest, is rephrase the question. Architecture is not a combinatorial system *tout court* any more than a language is a combinatorial system made up of words and rules of combination. In language, most – almost all in combinatorial terms – of the grammatically correct sequences of words of a language have no meaning, and are not in that sense legitimate sentences in the language. It is how (and why) these combinatoric possibilities are restricted that is the structure of the language. So with architecture. Most combinatorial possibilities are not buildings. The question is why not? How is the combinatorial field restricted and structured so as to give rise to the forms that exist and others that might legitimately exist? It is this that will be the theory of architectural form – the laws that restrict and structure the field of possibility, not the combinatorial laws of possibility themselves.

How then should we seek to understand these restrictions that structure the field of architectural possibility? There are a number of important clues. First, as the results reported in Chapter 4–8 show, the configurational properties of space, that is of the p-complex, are the most powerful links between the forms of built environments and how they function. It is a reasonable conjecture from these results, and their generality, that, in the evolution of the forms of buildings, factors affected the p-complex may dominate those affecting the a-complex. Bodily factors affecting the a-complex may create certain limits within which p-complexes evolve, but buildings are eventually structured by factors which affect the evolution of the p-complex, because it is the p-complex that relates to the functional differences between kinds of buildings.

Second, the properties of p-complexes that influence and are influenced by function tend to be global, or at least globally related, configurational properties, such as integration, that is, properties which reflect the relations of each space to many, even all, others. For example, the average quantity of movement along a particular line is determined not so much by the local properties of that space through which the line passes considered as an element in isolation, but by how that line is positioned in relation to the global pattern of space created by the street system of which it is a part (see Chapter 4). In general we may say that configuration takes priority over the intrinsic properties of the spatial element in relating form to function.

These conclusions may be drawn as generalisations from the study of a range of different types of building and settlement. However, there is a further, more general, conclusion that may be drawn from these studies which has a direct

Is architecture an ars combinatoria?

and powerful bearing on our present concerns. If we consider the range of cases studied as instances of real p-complexes within the total realm of the possible, we find that as complexes become larger they occupy a smaller and smaller part of the total range of possibility from the point of view of the total spatial integration of the complex, crowding more and more at the integrating end of possibility as complexes grow. For example, the recent doctoral study of over 500 English houses from the mid ninteenth to early twentieth century already referred to[6] with a mean size of 23.6 cells, has found most of the houses lie within the most integrating 30 per cent of the range of possibility and all within 50 per cent. Analysis of large numbers of buildings over a number of years suggest that at around 150 cells, virtually all buildings will be within the shallowest 20 per cent of the range of possibility, and most much below it, at 300 cells, nearly all will be within the bottom 10 per cent, and at around 500 most will be within the bottom 5 per cent. It is clear that as buildings grow, they use less and less of the range of possible p-complexes. The same is true of axial maps of settlements.[7]

In short, the most significant properties of p-complexes seem to be related to the degree and distribution of spatial integration – that is, the topological depth of each space from all others – in the complex. It follows that if we can understand theoretically how these characteristic properties of integration are created, then we will have made some significant progress towards understanding how architectural possibility becomes architectural actuality. How then does integration arise in a p-complex in different degrees and with different distributions? The simple fact is that the properties of any p-complex, however large, are constructed only by way of a large number of localised physical decisions: the placing of partitions, the opening of doors, the alignment of boundaries, and so on. What we need to understand in the first instance is how the global configurational properties of p-complexes space are affected by these various types of local physical change. It will turn out that the critical matter is that every local physical move in architecture has well-defined global spatial effects in the p-complex, including effects on the pattern and quantity of integration. It is the systematic nature of these effects by which local physical moves lead to global spatial effects that are the key to how combinatorial possibility in architecture is restricted to the architecturally probable, since these are in effect the laws by which the pattern and degree of integration in a complex is constructed.

Once we understand the systematic nature of these laws, we will be led to doubt the usefulness, and even the validity of the combinatorial theory of architecture in two quite fundamental ways. First, we will doubt the usefulness of the idea of spatial 'elements', because each apparent spatial element acquires its most significant properties from its configurational relations rather than from its intrinsic properties. Even apparently intrinsic properties such as size, shape and degree of boundedness will be shown to be fundamentally configurational properties with global implications for the p-complex as a whole. In effect, we will find that configuration is dominant over the element to the point where we must conclude that the idea of an element is more misleading than it is useful.[8] Spatial elements, we will show, are properly seen not as free-standing 'elements', with intrinsic properties, waiting to be brought into combination with others to create complexes of such properties, but as local spatial strategies to create global configurational effects according to well-defined laws by which local moves induce global changes in spatial configurations.

The second source of doubt will follow from the first: it is not combinatorics *per se* which create complexes but the local to global laws which restrict combinatorics from the vast field of architectural possibility to certain well-defined pathways of architectural probability. The theory we are seeking lies not in understanding either the theoretically possible or the real in isolation, but in understanding how the theoretically possible becomes the real. We will suggest that the passage from possibility to actuality is governed by laws of a very specific kind, namely laws which govern the relation between spatial configuration and what I will call 'generic function'. Generic function refers not to the different activities that people carry out in buildings or the different functional programmes that building of different kinds accommodate, but to aspects of human occupancy of buildings that are prior to any of these: that to occupy space means to be aware of the relationships of space to others, that to occupy a building means to move about in it, and to move about in a building depends on being able to retain an intelligible picture of it. Intelligibility and functionality defined as formal properties of spatial complexes are the key 'generic functions', and as such the key structures which restrict the field of combinatorial possibility and give rise to the architecturally real.

The construction of integration

Let us begin with figure 8.1f, a 6×6 half-partitioned a-complex with an isomorphic p-complex inscribed within it, that is, all partitions are permeable. What we are interested in is how the key global configurational property of integration is affected by closing and opening the central sections of the partitions. To make the process as transparent as possible, instead of using i-values, we will use the total depth counts from each cell from which the i-value is calculated. Half-partitions may be turned into full partitions by adding 'bars', in which case the cells either side become separated from each other, without direct connection. Half-partitions can also be eliminated, in which case the two cells become a single space. If all partitions to a cell are barred, then that cell becomes a block in the system.

Now as we already know from the analysis of shape in Chapter 3, the p-complex of figure 8.1f will already have a distribution of i-values, which we can show in figure 8.2a as total depth values, that is, the total depth of each cell from all the others, with the sum, 5040, shown below the figure. It is important for our analysis that we understand exactly how these differences arise, since all is not quite as it seems. We will, it turns out, need to make a distinction between the shape of the complex and the boundary of the complex. At first sight, it is clear that the differences between the cells are due to the relation of the cell to the boundary of the complex. Corner cells have most depth, centre edge rather less, then less towards the centre. If we change the shape of the aggregate, say into a 12 x 3 rectangle, as in figure 8.2b then all the individual cell total depths will change, as will the total depth for the

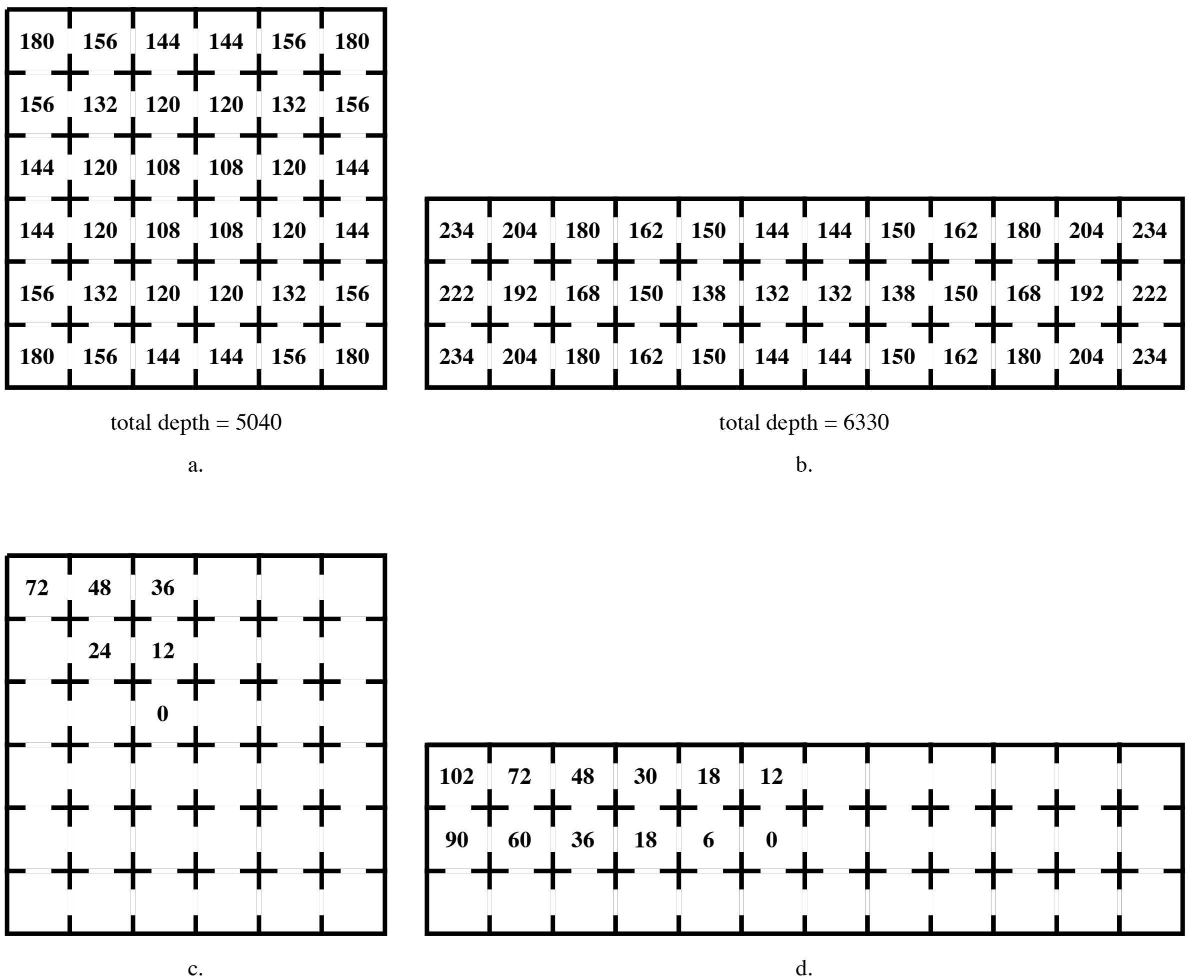


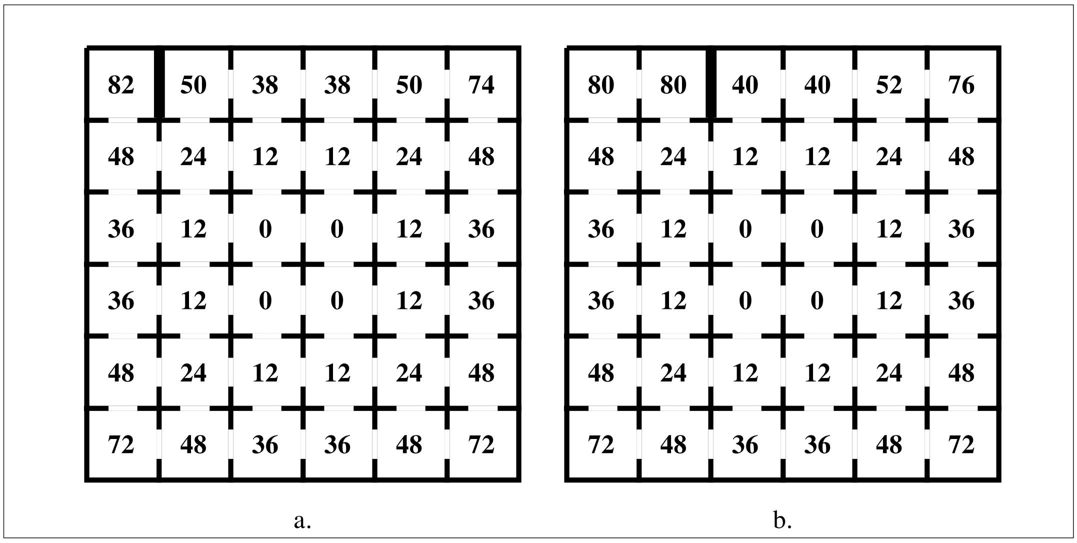
Figure 8.2

aggregate as a whole (6330) reflecting the changing relations of cells to the boundary. However, if we eliminate the boundary by wrapping either of the two aggregates first round a cylinder so that left joined to right, and then into a torus so that top joined to bottom, then the total depths for all cells in each aggregate would be the same, since starting from each and counting outwards until we have covered all the cells, we will never encounter a boundary and so will find the same pattern of depth from each cell. The total depths of the cells would in fact be equal to the minimum depth of the cells in the bounded aggregate, that is the group of four at the centre of the square form, whose value is 108, and the pair at the centre of the rectangular form, whose depth is 132. However, this implies that in spite of the removal of the boundaries, these differences between the square and rectangular shapes still survive. These differences in total depth values are it seems the product of the shape of the aggregate but not of its boundary.

This can be demonstrated by a simple thought experiment. Take a cellular aggregate, say the six by six square and wrap it onto a torus, thus removing the boundary. Select any 'root' cell and construct a justified graph – that is a graph in which levels of depth of nodes from an initial node are aligned above a selected root node in a series of layers representing depth – in which all cells sharing a doorway with the root are the first layer, all those sharing a doorway to a first layer cell are the second layer, and so on. When the graph reaches any cell adjacent to the boundary in the original bounded aggregate in the plane, any next, deeper cell with which a cell in the justified graph shares a doorway will already be in another branch of the graph. Thus the justified graph finds the limits of the original shape of the aggregate, even though the boundary has been eliminated by wrapping on the torus.

It follows that the uniform depth value that will be found in any shape on a torus will reflect the shape and will be equal to the minimum depth of the original aggregate in the plane. This will be 108 for the square form and 132 for the rectangle. A depth of 108 per cell (three times the number of cells in the complex) can therefore be said to be the depth due to the square form having a square shape and 132 the depth due to the rectangular form having a rectangular shape. When dealing with a standard shape therefore we may, if we wish, eliminate this amount of depth from each cell, and deal only with the depth due to the boundary. These remaining depths are shown for the 6 x 6 square and the 12×3 rectangle in figures 8.2c and d. These boundary related depths are due to the fact that the aggregate boundary is barred from its surrounding region. If we were to open all cells to the outside by opening the boundary, and treating the outside region as an element in the system to be included in depth calculations, then clearly the depth values would all change, particularly if we counted the outer region as a single space, in which case cells close to the boundary would have less depth than cells at the centre. This alerts us to the fact that in considering the barring – that is the conversion of half partitions into full partitions – in a cellular aggregate, the boundary is itself an initial partitioning, and like any other partitioning it has effects on the distribution of

Figure 8.3



depth in the aggregate. Bearing this in mind, we may now return to the plane, and hold shape and boundary steady by considering only the square form, in order to explore the depth effects of adding further barrings within the aggregate.

Is architecture an ars combinatoria?

It is obvious that further internal barring will increase the total depth for at least some cells, since it will have the effect of making certain trips from cell to cell longer. It is perhaps less obvious that the quantity, as well as the distribution, of extra depth created by bars will vary with the location of the bar in relation to the boundary. For example, if we place a bar in the leftmost horizontal location in the top line of cells in figure 8.1, as in figure 8.3a, the total depth in the aggregate will be increased from 5040 to 5060, an additional 20 steps of depth, while if we place the bar one to the right, as in figure 8.3b, then the increase in total depth will be from 5040 to 5072, an additional 32 steps.

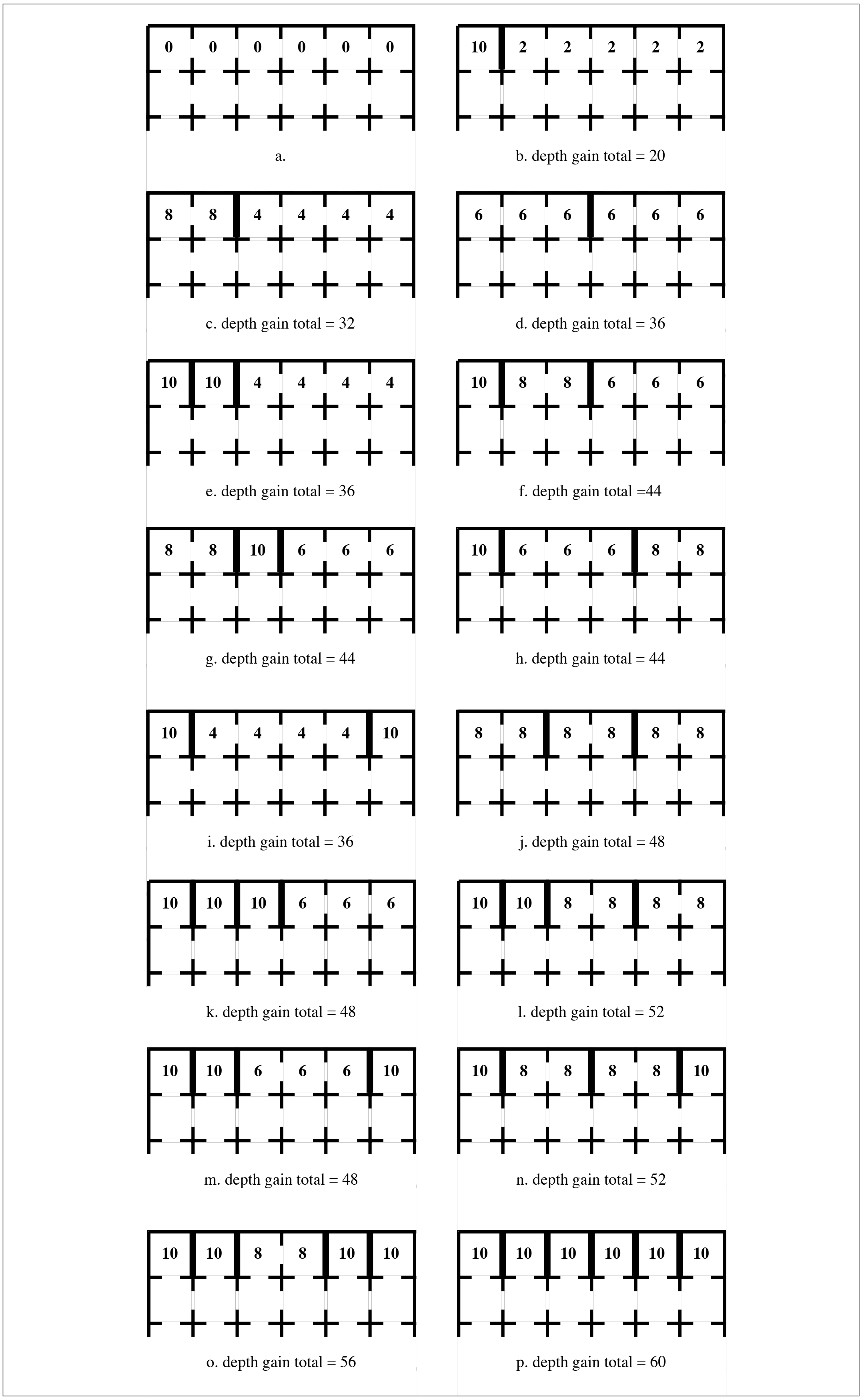
How does this happen? First, all the 'depth gain' in figures 8.3a and b is on the line in which the bar is located. On reflection, this must be the case. Depth gain happens when a shortest route from one cell to another requires a detour to an adjacent line. Evidently, any other destination on that adjacent line or on any other line will not require any modification to the shortest path, unless that line is itself barred. Depth gain for single bar must then be confined to the line on which the bar occurs. But placing the bar at different points on the line changes the pattern of depth gain for the cells along the line. Each cell gains depth equal to twice the number of cells from which it is linearly barred, because each trip from a cell to such cells requires a two-cell detour via an adjacent line. Evidently this will be two way, and the sum of depths on the two sides of a single bar will thus always be the same. It follows that the depth gain values of individual cells will become more similar to each other as the bar moves from edge to centre, becoming identical when the bar is central. It also follows that the total depth gain from a bar will be maximised when the bar is at or near the centre of the line, and will be minimised at the edge. This is illustrated for edge to centre bars on a 6-cell line in figure 8.4 a, b c, and d.

The fact that an edge location for a partition minimises depth gain but maximises the differences between cells, while a central location maximises depth gain but minimises differences, is a highly significant property. It means that decisions about where to place a bar, or block a doorway, have implications for the system beyond the immediate region of the bar. If we define a 'local physical decision' as a decision about a particular bar within a system, and a 'global spatial effect' as the outcome of that decision for the system as a whole, it is clear that local decisions do have quite systematic global effects. In these cases, the systematic effects follow what we might call the 'principle of centrality'.

It might be useful to think of such 'local-to-global' effects as 'design principles', that is, as rules from which we can forecast the global effect of a local barring decision by recognising what kind of barring we are making. In this case the design principles are two: that the depth gain from a bar is minimised when the bar is placed at the edge and maximised when placed at the centre; and that edge bars make for greater depth gain differences between some cells and others, while central partitions equalise depth gain.

Similar principles govern local-to-global effects when we add a second bar in different locations as in figure 8.4e-j. Depth gains for each cell are equal to

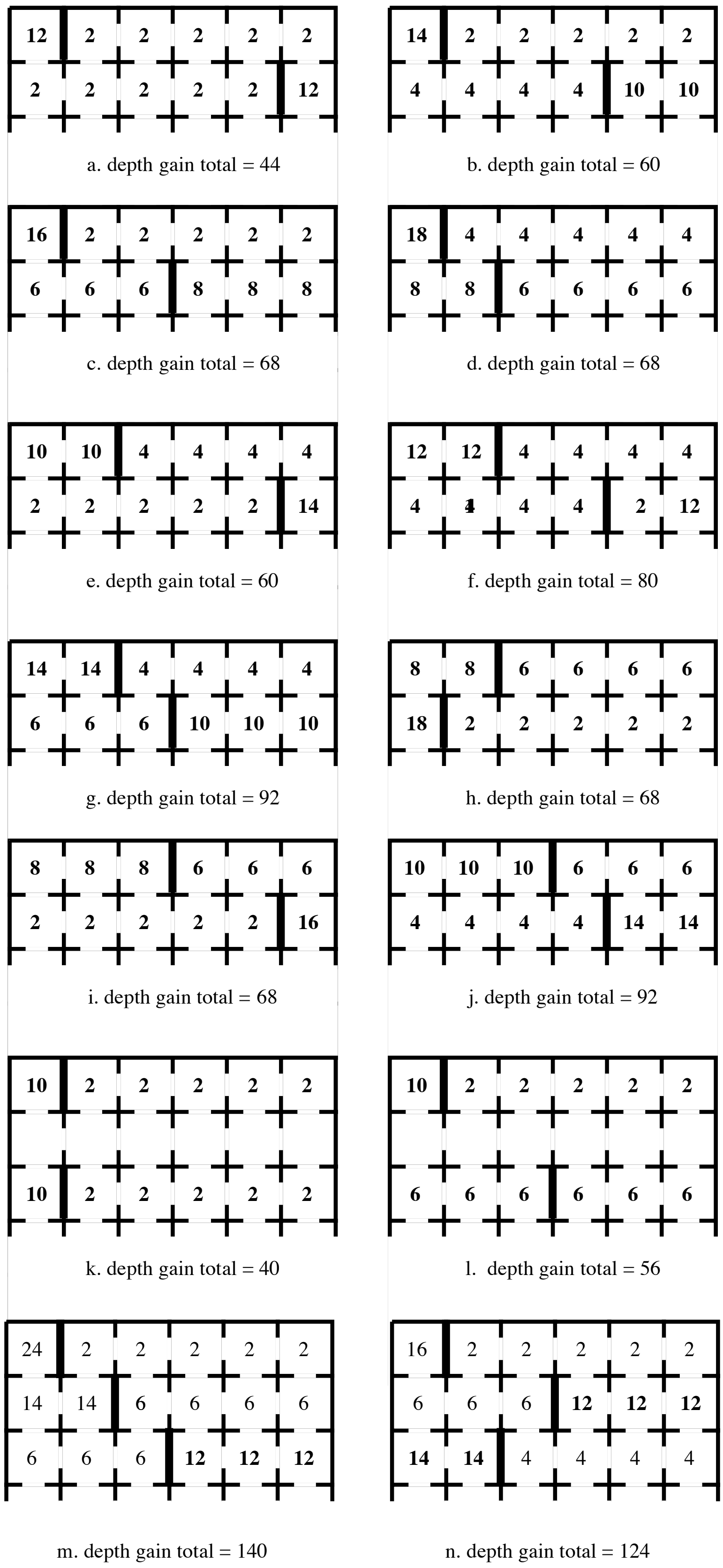
Figure 8.4


0 0 0 0 0 0
a.
10 2 2 2 2 2
b. depth gain total = 20
8 8 4 4 4 4
c. depth gain total = 32
6 6 6 6 6 6
d. depth gain total = 36
10 10 4 4 4 4
e. depth gain total = 36
10 8 8 6 6 6
f. depth gain total =44
8 8 10 6 6 6
g. depth gain total = 44
10 6 6 6 8 8
h. depth gain total = 44
10 4 4 4 4 10
i. depth gain total = 36
8 8 8 8 8 8
j. depth gain total = 48
10 10 10 6 6 6
k. depth gain total = 48
10 10 8 8 8 8
l. depth gain total = 52
10 10 6 6 6 10
m. depth gain total = 48
10 8 8 8 8 10
n. depth gain total = 52
10 10 8 8 10 10
o. depth gain total = 56
10 10 10 10 10 10
p. depth gain total = 60

twice the number of cells on the far side of the nearest bar. For each cell, bars other than the nearest on either side do not affect depth since once a detour to an adjacent line has been made, then it can be continued without further detour to reach other cells on the original line, provided of course there is no bar on the adjacent line (see below). Figure 8.4k-p then shows that depth effects of three to five bars are governed in the same way, ending with the fully barred line in which each cell gains depth equal to twice the number of other cells in the line. These examples illustrate a second principle: that once a line is barred, then depth gain from the next bar will be minimised by placing it within the shortest remaining line of cells, and maximised by placing it in the longest. We can call this the 'principle of extension': barring longer lines creates more depth gain than barring shorter lines. Within each line, of course, the principle of centrality continues to hold, and the distribution of depth gains in the various cases in figure 8.4 follow these both in the principle of extension and the principle of centrality. Thus taking figures 8.4g and j, each has a bar in the second position in from the left, but g then has its second bar immediately adjacent in the third position in from the left, while j has its second bar two positions away, equidistant from the right boundary of the complex. This is why g has less depth gain than j in spite of its second bar being in a more central location in the complex as a whole, because, given the first bar, what counts is the position of the next bar in the longest remaining lines, and in j the bar is placed centrally on that line. This shows an important implication of the principles of centrality and extension: when applied together to maximise depth gain, they generate an even distribution of bars, in which each bar is as far as possible from all others; while if applied to minimise depth gain, bars becomes clustered as close as possible to each other along lines.

Suppose now that instead of locating the second bar on the same line we locate it on an adjacent line. Figure 8.5a-j shows the sequence of possibilities for the location of the second bar, omitting, for the time being (but see below) the case where we join bars contiguously in a line. When barred lines are adjacent, then for each line, the depth gain is greater than for each bar alone, but the effect disappears when the two barred lines are not adjacent, as in the final two cases, k and l. The effect is identical if the two bars are on adjacent lines away from the edge. These effects are best accounted for by seeing each barring of two adjacent lines as dividing the pair of lines into an 'inner zone', where there is only one bar to circumvent in each direction, and two 'outer zones' from which two bars must be circumvented to go from one to the other. The conjoint effect is entirely due to the outer zones, in that to go from one outer zone to the other, there is a further bar to circumvent once a detour to the adjacent line is taken to circumvent the first bar. Depth gain for a cell is therefore equal to twice the number of cells that lie beyond bars on either line. Thus the value of twelve in the leftmost example in the top row is the product of twice the five cells on the far side of the bar in the top row, plus twice the single cell on the far side once you move from the top to the second row. Similarly, the total depth of two for each of the cells to the right of the bar in the top

Figure 8.5


12
2
2
2
2
2
2
2
2
2
2
12
a. depth gain total = 44
14
2
2
2
2
2
4
4
4
4
10
10
b. depth gain total = 60
16
2
2
2
2
2
6
6
6
8
8
8
c. depth gain total = 68
18
4
4
4
4
4
8
8
6
6
6
6
d. depth gain total = 68
10
10
4
4
4
4
2
2
2
2
2
14
e. depth gain total = 60
12
12
4
4
4
4
4
4
4
4
2
12
f. depth gain total = 80
14
14
4
4
4
4
6
6
6
10
10
10
g. depth gain total = 92
8
8
6
6
6
6
18
2
2
2
2
2
h. depth gain total = 68
8
8
8
6
6
6
2
2
2
2
2
16
i. depth gain total = 68
10
10
10
6
6
6
4
4
4
4
14
14
j. depth gain total = 92
10
2
2
2
2
2
10
2
2
2
2
2
k. depth gain total = 40
10
2
2
2
2
2
6
6
6
6
6
6
l. depth gain total = 56
24
2
2
2
2
2
14
14
6
6
6
6
6
6
6
12
12
12
m. depth gain total = 140
16
2
2
2
2
2
6
6
6
12
12
12
14
14
4
4
4
4
n. depth gain total = 124

Is architecture an ars combinatoria?

row reflects the fact that only one cell is on the far side of the bar in the top row, and none are in the second row. This calculation of depth gain will work for any number of rows of cells, providing that the bars are non-contiguous. Non-contiguity of bars means that there is always a 'way through' for a shortest path.

If we then add a third (non-contiguous) bar on a third line, then there are two alternative possibilities. If the three bars are in echelon, as in figure 8.5m, then 'outer zone' cells on all three lines will gain depth additively equal to twice the number of cells in all the opposite outer zones. This is because when the bars are in echelon, then every detour to an adjacent barred line means that the bar on that line is still beyond where you are on that line, so a further detour is necessary. Inner zone cells gain only twice the number of cells in the outer zones of their own lines.

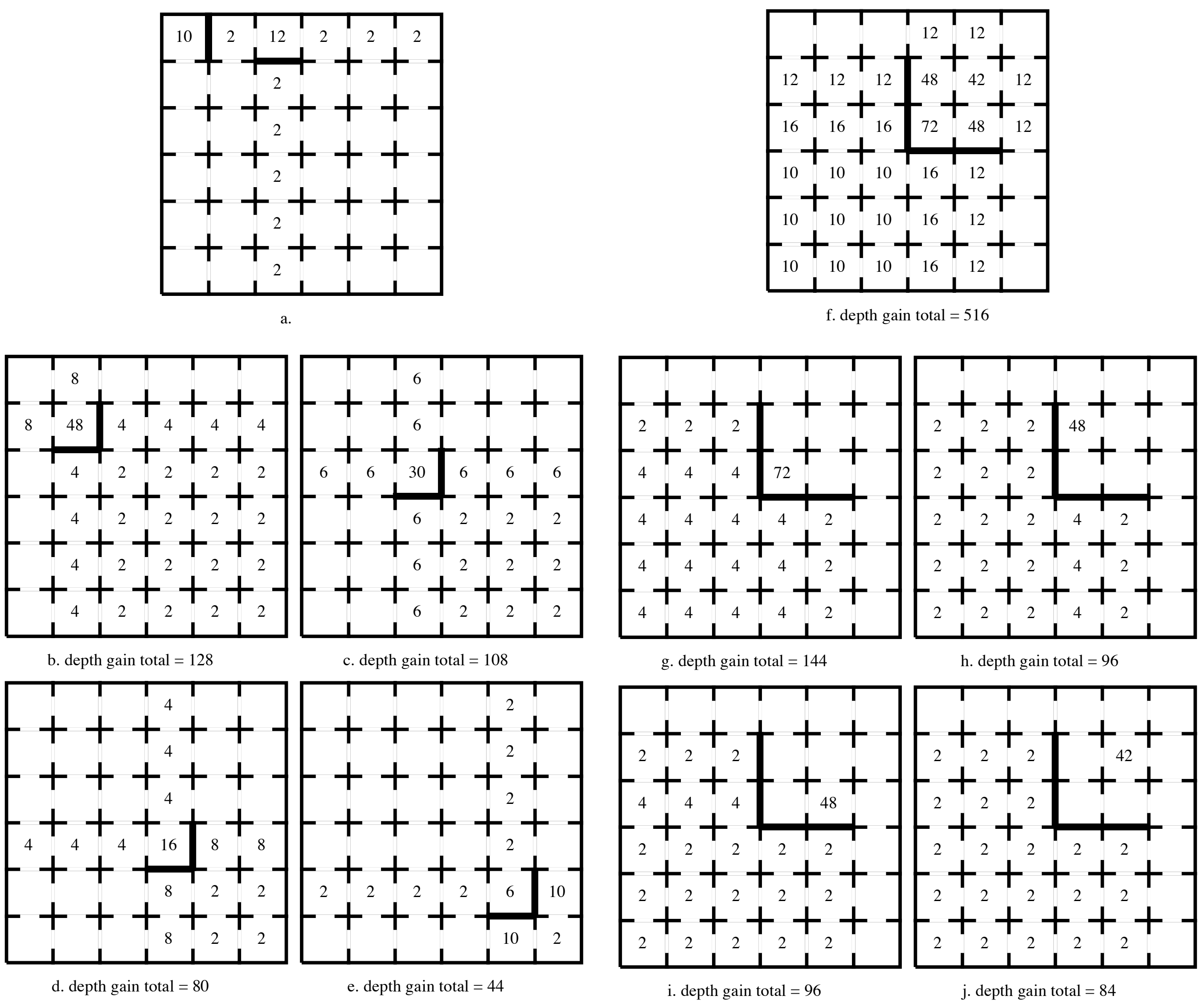
If the bars are not in echelon, as in figure 8.5n, then the gain will only be as from a pair of adjacent lines since the bar on the central line must be so placed as to allow a 'way through'. The central line will, however, gain depth from its relation to both adjacent lines, and can be counted first in a pair with one, then with the other. If four non-contiguous bars are on four adjacent lines, then the depth gain is according to whether trios of lines are in echelon or not, and so on. If there are two or more bars on the same line, then the calculations will be according to the formula already outlined. If one of the adjacent lines is an edge line, then likewise, this can be calculated according to the formula already explained.

These are the possible non-contiguous barrings on the same general alignment (i.e in this case all are horizontal). What about the addition of a second (or more) non-contiguous bar on the orthogonal alignment, as in figure 8.6a? We already know the effect of the second bar on its own line. Does it have an effect on the line of the first bar? The answer is that is does not and cannot, provided it is non-contiguous, because while it is non-contiguous there will always be a 'way through' for shortest paths from cells on other alignments. Depth gain resulting from a bar on a certain alignment can never be increased by a bar orthogonal to that alignment, while the bars are non-contiguous.

What then are the effects of contiguous bars? There are two kinds: linearly contiguous bars, in which two or more partitions form a single continuous line; and orthogonally contiguous bars, in which two or more bars form a right-angle connection. Within each we can distinguish contiguous bars which link with another bar at one end, and those which link at both ends. First let us look at the right angle, or L-shaped, case for the single connected bar. Figures 8.6b–e show the depth gain pattern for the simplest case, a two bar L-shape, located at four different positions. The first thing we note is that in all cases the depth gains on 'either side' of the L are in total equal, though very differently distributed. In 8.6b, where the L faces into the top left corner, the depth gain forms a very high peak within the L, which is made up of two elements: first, the depth gains along each of the lines of cells partitioned by the bar, of the kind we have seen already; and second by the conjoint effect of the two bars forming the L, in creating a 'shadow' of cells, each with a depth gain of 2, which mirror the L shape on the outside diagonal to it. This is a phenomenon we have not see before, since with non-contiguous bars all depth gains can be accounted for by the effects of individual bars.

As the L-shaped bar is moved from top left towards the bottom right, while maintaining its orientation, as in 8.6c,d and e, we find that although the individual effects of each of the constituent bars making up the L remains consistent with the effects so far noted, the conjoint 'shadow' effect diminishes, because there is less and less scope for the 'shadow' as the L moves towards the bottom right and the L shape follows, rather than inverts, the L formed by the corner of the outer boundary. We see then that in this case the effect of moving the L from the centre towards the corner will be to diminish depth gain, as expected, as the L moves towards a corner from which the L faces outwards, but to increase it as the L moves towards

Figure 8.6



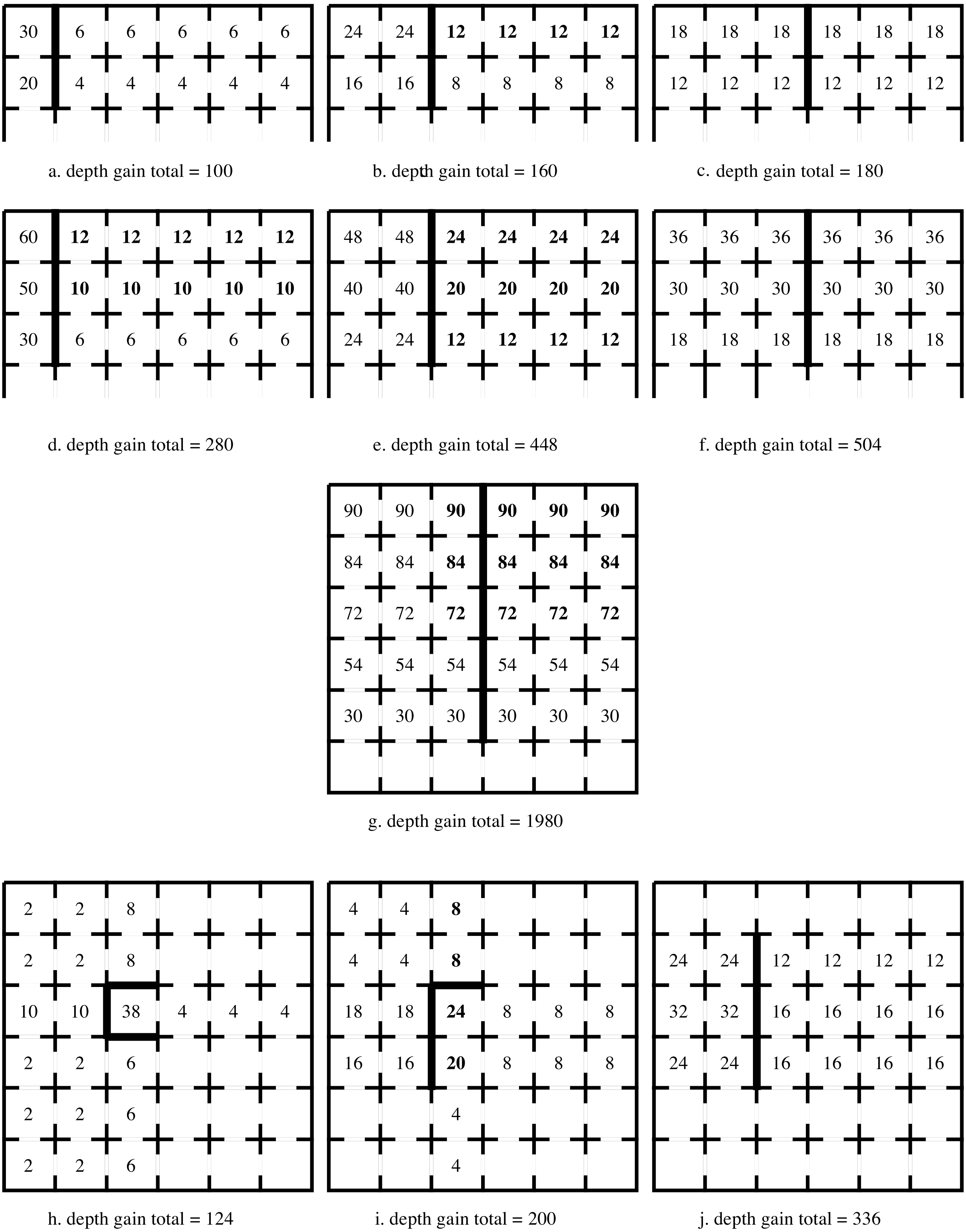
a corner where the L faces inwards towards the corner.

At first sight, this seems to contradict the principle that edge partitions cause less depth gain and central partitions more. In fact, what we have is a stronger instance of the effect noted in figure 8.4a, where the most peripherally located partition created the least depth gain overall but the greatest depth gain for the single cell. The depth gain was focused, as it were, in a single cell. In 8.6b, the depth gain is even more powerfully focussed in a single cell, both because it focusses both the gain from the two bars making up the L, but also from the 'shadow'. In other words what counts as the 'other side' of the partition is expanded by forming contiguous partition into an 'enclosure'. Enclosure, we might say, means 'enclosure with respect to what'. The greater the area 'with respect to which' an 'inside' region is enclosed, then the greater the enclosure effect by the focussing of depth gain. This is, in effect, a generalisation of the 'principle of extension' by which greater overall depth gain arises from the greater scope of the effect of the partition. In figure 8.6b, this extension on the 'other side' of the enclosure includes the area between the two alignments affected by the partition, and this increases its extension.

This effect will increase if we add new contiguous bars to the original L-shape. Figure 8.6f for example shows the depth gain pattern for an L-shape whose arms are twice as long as in the previous figure. The depth gain pattern is similar to that for single L-shapes, but even more extreme. Figures 8.6g–j break this down by taking each of the cells on the open side of the barring and showing the shadow due to that cell. This is calculated by taking each open side cell in turn and calculating the detour value for each shadow cell. The shadow shown in figure 8.6f evidently, is the sum of these sub-shadows of figure 8.6g–k, plus those of the four cells on the 'open' side of the L (which are not shown).

Next consider the linear contiguity of bars. Figure 8.7a–g shows a series of cases in which bars are first extended linearly to double unit length and moved across from edge to centre, and then triple unit length. Depth gains are larger even than for L-shaped bars, and the rate of gain increases, not only as the line of bars is moved from edge to centre, but also, even more dramatically, as the number of bars formed into a continuous line is increased. For example, the depth gain from a single edge bar is 20, rising to 36 as the bar moves to the centre, but if we expand the bar linearly to a pair, the gain is 180 and if we add a third then the gain is 504. This reflects a simple fact that to detour round one bar – say an edge bar – to a cell that was initially adjacent requires a 2 cell detour. However, if a second bar is added in line, then the detour will be 5 cells, and if a third is added, the detour will be 7 cells, and so on. The contiguous line of bars is the most effective way of increasing depth in the system, first because it is the most economical way of constructing an object requiring the longest detour from cells on either side to the other and second because the longer the bar the more it has the effect of increasing the number of cells on either side of it, that is, it has the effect of barring the whole aggregate. Evidently, this 'whole object barring' will have more depth gain to the degree that the

Figure 8.7


30 6 6 6 6 6
20 4 4 4 4 4
a. depth gain total = 100
24 24 12 12 12 12
16 16 8 8 8 8
b. depth gain total = 160
18 18 18 18 18 18
12 12 12 12 12 12
c. depth gain total = 180
60 12 12 12 12 12
50 10 10 10 10 10
30 6 6 6 6 6
d. depth gain total = 280
48 48 24 24 24 24
40 40 20 20 20 20
24 24 12 12 12 12
e. depth gain total = 448
36 36 36 36 36 36
30 30 30 30 30 30
18 18 18 18 18 18
f. depth gain total = 504
90 90 90 90 90 90
84 84 84 84 84 84
72 72 72 72 72 72
54 54 54 54 54 54
30 30 30 30 30 30
g. depth gain total = 1980
2 2 8
2 2 8
10 10 38 4 4 4
2 2 6
2 2 6
2 2 6
h. depth gain total = 124
4 4 8
4 4 8
18 18 24 8 8 8
16 16 20 8 8 8
4
4
i. depth gain total = 200
24 24 12 12 12 12
32 32 16 16 16 16
24 24 16 16 16 16
j. depth gain total = 336

object is barred into two equal numbers of cells. Thus in figure 8.7g the long central bar comes as close as possible to dividing the whole object into two equal parts. Figure 8.7h–j then demonstrates the effect of linearity on three contiguous bars. In all three, at least two bars are located in the second position from the edge. In 8.7h, the bars are formed into a U-shape giving a total depth gain of 124, 28 more than would be gained by the lines independently if they were non-contiguous, and with a very strong peak inside the enclosure. In figure 8.7i, which is a three-bar L-shape, the total depth gain is 200, 104 more than the lines would have independently, and with a less strong peak within the enclosure. In 8.7j, the total gain is 336, 240 more than for the lines independently, and with a much more even spread of values, without any single peak. These differences thus arise simply from the shape formed by the three contiguous lines. The principle is that the more we coil up bars, and create a concentrated peak of depth gain within the coiled up bars, then the less the overall depth gain. Depth gain in the whole system is maximised when bars are maximally uncoiled and construct a maximally linear 'island' of bars. Since the U-shape of 8.7h approximates a 'room', we can say that the most integration efficient way of arranging three contiguous bars is to form them into 'rooms'. Such 'rooms' will not only have the least depth gain effect on the spatial complex, but will also maximise the difference between the depth-gain of a single space (i.e. the 'room') and that in the other spaces of the system. This is the phenomenon we first noted for edge partitions in figure 8.4.

Now if we reflect on figure 8.7j, we can see that all the depth gain apart from that due to the individual bars is to the central bar and to the fact that it connects two ways to form the line of three. This means that if we start from a situation in which we have the two outer bars, then the addition of the single bar connecting the two outer bars into a line in itself adds a depth gain of 272. This double connecting of bars to form a line is the most powerful possible move in creating additional depth, not least because it must necessarily have the effect of eliminating a ring from the system.

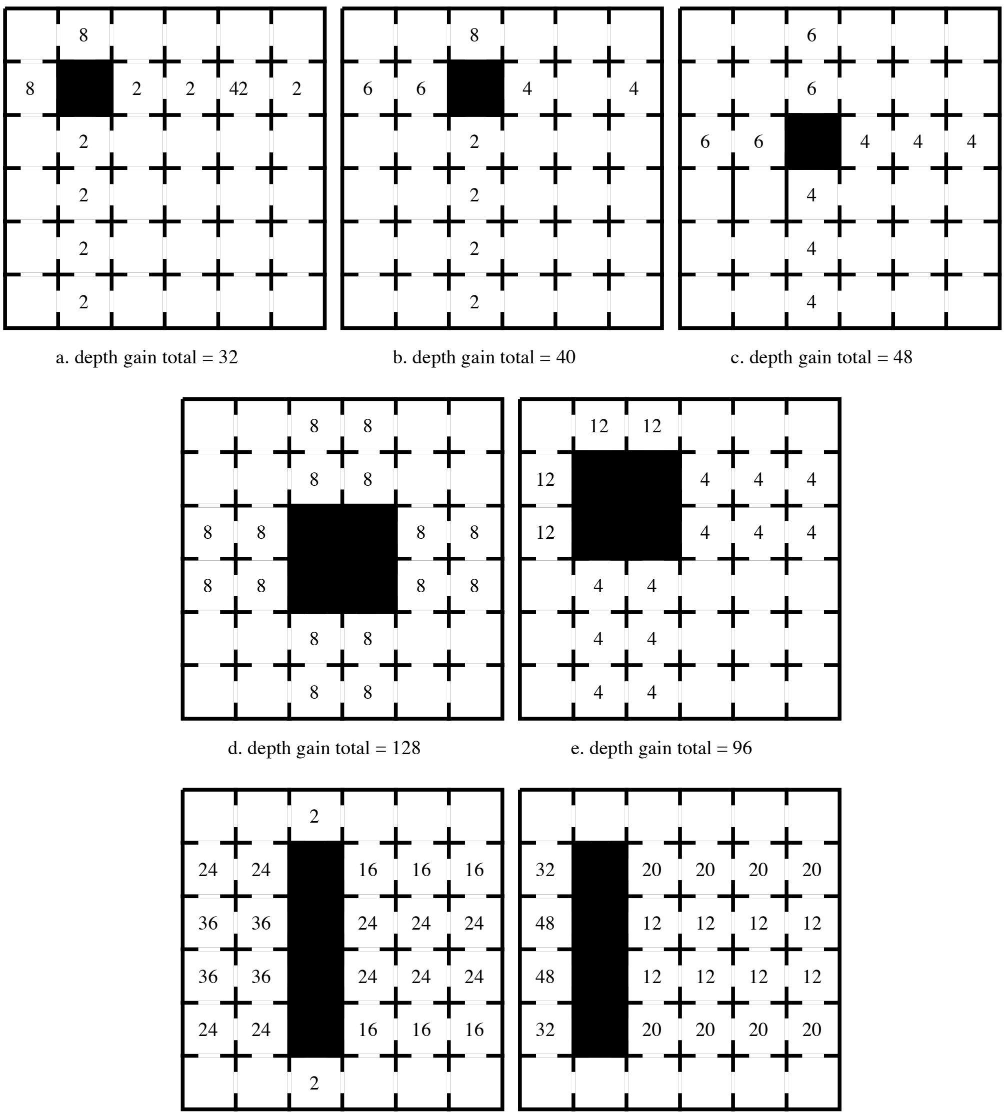
We may summarise all these effects in terms of four broad principles governing the depth gain effects of bars: the principle of *centrality:* more centrally placed bars create more depth gain than peripherally placed bars; the principle of *extension*: the more extended the system by which we define centrality (i.e. the length of lines orthogonal to the bar) then the greater the depth gain from the bar; the principle of *contiguity:* contiguous bars create more depth gain than non-contiguous bars or blocks; and the principle of *linearity:* linearly arranged contiguous bars create more depth gain than coiled or partially coiled bars. All four principles govern local-to-global effects in that each individual local physical move has quite specific global effects on the spatial configuration as a whole. At the same time these effects are dependent on the number and disposition of bars and blocks that already exist in the system. The four principles allow us to keep track of the complex inter-relationships between what is already in the system and the global consequences of new moves. We may therefore expect to be able to construct processes in which different sequences of barring moves will give rise to different global configurational properties.

Elementary objects as configurational strategies

We will see shortly that this is the case. But first we must show that the same principles that govern the opening and closing of partitions, also govern all other types of spatial moves which affect integration such as the creating of corridors, courts or wells, and even changes in the shape of the envelope of the complex. Let us first consider wells. Wells are zones within a complex which are inaccessible from the complex and therefore not part of the spatial structure of the complex. They act in effect as blocks in the system of permeability. We will see that the effects of blocks of different shapes and in different locations have configurational effects on the whole system which follow exactly the same principles as those for bars.

First, let us conceptualise blocks in terms of the barring system we have so far discussed. A block is an arrangement of bars we have so far disallowed, that is, an arrangement of four or more bars in such a way as to form a complete

Figure 8.8



a. depth gain total = 32

b. depth gain total = 40

c. depth gain total = 48

d. depth gain total = 128

e. depth gain total = 96

f. depth gain total =484

g. depth gain total = 420

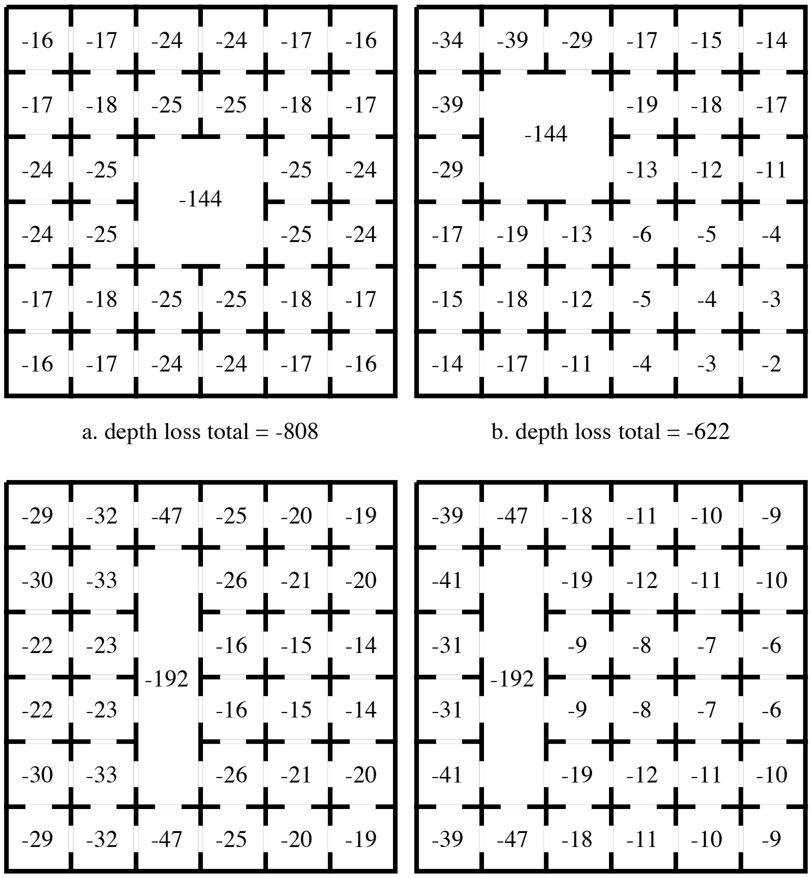
enclosure, so that one or more spaces is completely separated from the rest of the spatial system, and effectively eliminated from it. A block is in effect the elimination of one or more cells from the spatial system. Three possible cases of single cell elimination are shown in figure 8.8a, b and c with the resulting depth gains. Because the block bars lines in two directions all that happens is that the pattern of depth gain resulting from the blocks follows the edge to centre rules, as for bars. There will not, for example, also be 'shadow' effects, as with L-shaped bars, because the relation between the enclosed space and those on the other side of the L-shape, which created the 'shadow' has been eliminated by the complete closing off of the block. We must note of course that the depth gains figures are less than for a simple barring, but this is simply because one cell has been eliminated from the system. We may if we wish correct this by substituting i-values for depth gains, since these adjust depth according to the total number of cells in the system, but at this stage it is simpler to simply record the depth gains and note the effect of the elimination of a cell.

Figure 8.8d–g then shows four possible shapes and locations for blocks of four cells, together with the depth gains for each cell and the total depth gain indicated bottom right of the complex. As we would expect from the study of bars, the compact 2×2 block has much less depth gain than either of the linear 4×1 forms, and the linear forms have higher depth gains in central locations than peripheral locations (as would compact blocks). We may note that, as we may infer from bars, the depth gain effects from changes of shape are much greater than those from changes of location. But also of course the locational effects of high depth gain shapes – that is linear shapes – are much greater than the locational effects of low depth gain – or compact – shapes.

It is clear that in this way we can calculate the depth gain effect of any internal block of any shape and that it will always follow the general principles we have established for bars. However, there is another important consequence of this, namely that we can also make parallel calculation for blocks placed at the edge of the complex. The reason this is important is that such peripherally located blocks are not 'wells' which by definition are internal to the complex, but changes in the shape of the envelope of the complex. It is clear from this that we may treat changes in the external shape of the complex in exactly the same way as interior 'holes' within the complex. Since we have already shown that such 'holes' are special cases of barring, then there is a remarkable unification here. From the point of view of the construction of integration – which we already know to be the chief spatial correlate of function within the complex – it seems that partitions within the complex are the same kind of thing as changes to the shape of the complex, whether these are internal, as with wells, or external, as with changes in the envelope shape.

We will now show that the creation of larger spaces within a complex such as courts and corridors can also be brought within the scope of this synthesis and be shown to be the same kind of phenomenon and subject to the same laws. First, we must conceptualise what we mean by the creation of larger spaces in terms of

Figure 8.9



a. depth loss total = -808

b. depth loss total = -622

c. depth loss total = -1174

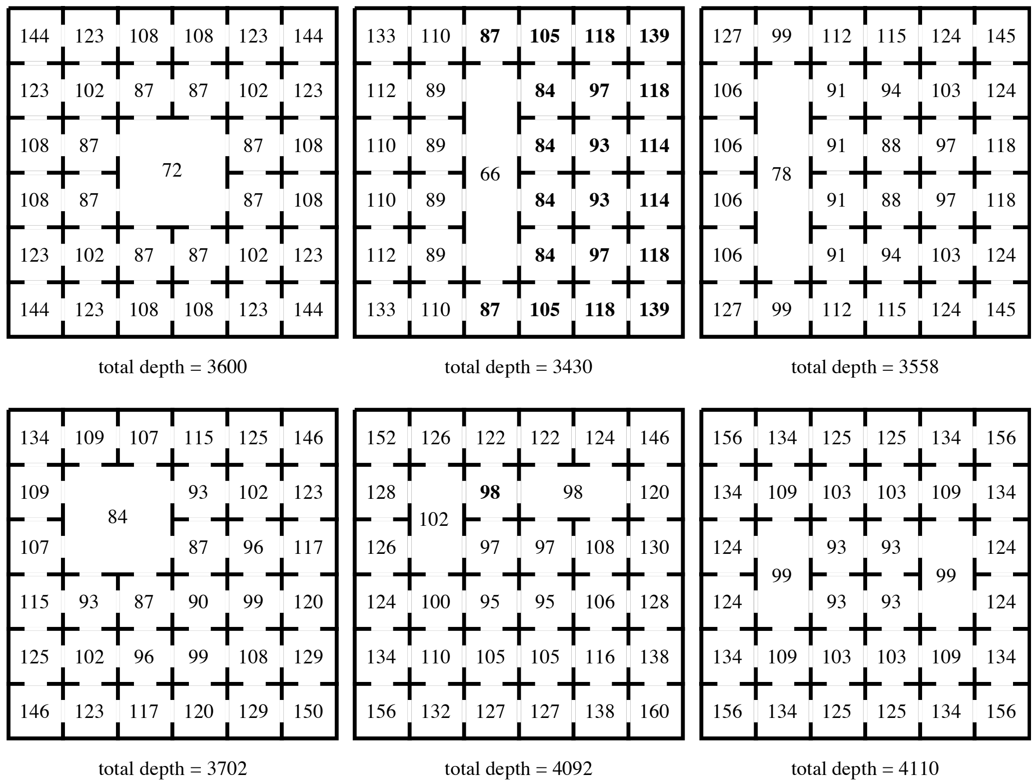
d. depth loss total = -1008

a barring process. Larger open spaces in the complex are created by eliminating the existing two-thirds partitions instead of completing the partition, and in effect turn two neighbouring spaces into what would then be identified as a single space. Figure 8.9a–d does this so as to substitute open spaces for the blocks shown in the previous cases, and gives the consequent depth loss (that is, integration gain) for each cell. The depth loss for the larger space is calculated by substituting the new value for the whole space for each of the values in the original form and adding them together. Total depth loss for each form is shown below the figure.

The first point to be noted is that the depth loss for a shape of a given size is a constant, regardless of location in the configuration. This is because from the point of view of the large space, the effect of substituting a single space for two or more spaces is to change the relations of those spaces with each other – that is to eliminate a certain number of steps of depth – but not to change the relations of those spaces to the larger system. However, although the depth loss for the larger space is constant, its effects on the rest of the system are not. In fact they vary in exactly the opposite way to the blocks. Whereas peripherally located blocks add less depth to the system than centrally placed blocks, peripherally placed open spaces eliminate less depth than centrally placed spaces; and a linear arrangement

Is architecture an ars combinatoria?

Figure 8.10



of cells into a single space has a greater depth loss (more integrating) effect than a square arrangement, and this effect is greater when the linear space is placed centrally than when it is placed peripherally.

The first four complexes of figure 8.10 show the same cases but marking each space with its total depth from the rest of the system rather than its depth loss. Here what we note is that identical larger spaces in different locations will have different total depths reflecting their location in the complex. It is only the depth loss from making two or more spaces into one that is identical, not the depth values of the location of these spaces in the complex. Thus we can see that a centrally placed open 'square' is more integrating (i.e. has less total depth) in itself than a peripherally placed one, and that a linear form will be more integrating than a compact form. These effects are of course exactly the inverse of those of blocks, and we may therefore say that they are governed by the same laws. In the two final examples in figure 8.10 the four open cells are arranged as two two-cell spaces rather than a single four-cell space and show another inverse principle: that contiguously joined spaces will always create more integration than a comparable number of discrete spaces.

Is architecture an ars combinatoria?

Thus the four principles of centrality, extension, contiguity and linearity which governed the depth gain effects of bars and blocks also govern the depth effects on the global system of creating larger open spaces, though in the contrary direction. More centrality for larger spaces means more integration, more extended lines from larger spaces means more integration, more continuity of larger spaces means more integration and more linearity of larger spaces means more integration. A useful bonus is that in the case of larger spaces we can actually see that the effects are not within the spaces themselves but are to do with the effect of the spaces on the remainder of the system.

We can now draw a significant conclusion. Not only partitions, internal walls and external shape changes but also rooms and larger linear or compact open spaces such as corridors and courts have all been shown to be describable in the same formal terms and therefore to be, in a useful sense, the same kind of thing. This has the important implication that we will always be able to calculate the effects of any spatial move in any system in a consistent way, and indeed to be able to predict its general effects from knowledge of principle. This allows us to move from a static analysis of the global implications of local changes in system to the study of dynamic spatial processes in which each local move seeks, for example, to maximise or minimise one or other type of outcome. When we do this we will find out that both the local configurations we call elements and the global patterns of the spatial complex as a whole are best seen as emergent phenomena from the consistent application of certain types of spatial move. We will call these dynamic experiments ‘barring processes’.

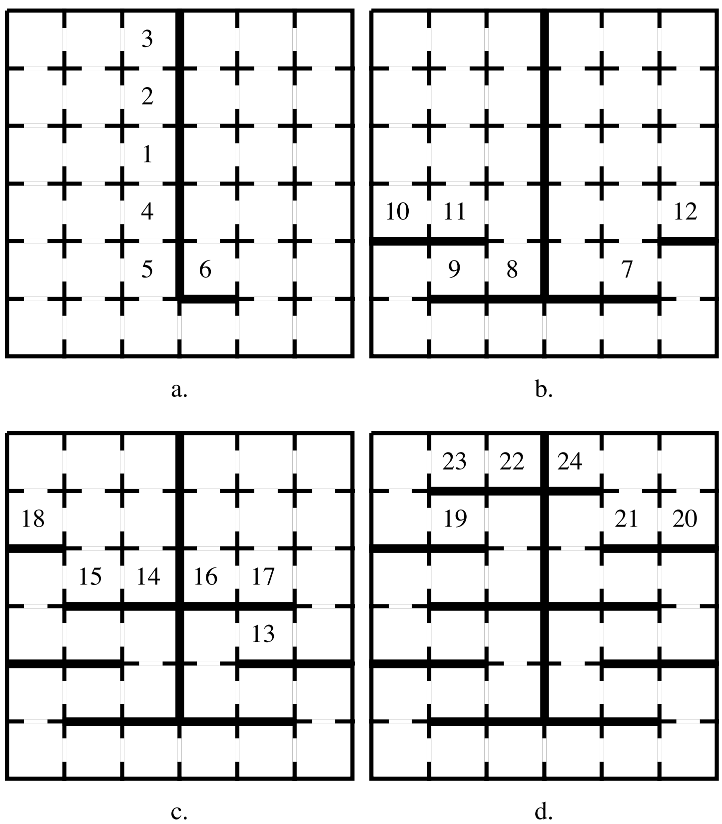
Barring processes

For example, we may explore barring processes which operate in a consistent way, say to maximise or minimise depth gain, and see what kind of cellular configurations result. In making these experimental simulations, it is clear that we are not imagining that we are simulating a process of building that could ever have occurred. It is unrealistic to imagine that a builder would know in advance the depth gain consequences of different types of barring. However, it is entirely possible that within a building tradition, a series of experiments in creating cellular arrangements would lead to a form of learning of exactly the kind we are interested in: that certain types of local move will have global consequences for the pattern as a whole which are either functionally beneficial or not. We may then imagine that our experiments are concerned not with simulating a one-off process of building a particular building, but of trying to capture the evolutionary logic of a trial-and-error process of gradually learning the global consequences of different types of local barring moves. In this sense, our experiments are about how design principles might be learnt rather than how particular buildings might be built.

First some definitions. We define a barring move as the placing of a single bar whose only known (or, on the evolutionary scale, discovered) consequence is its depth gain for the system as a whole. A barring manoeuvre is then a planned series of two or more moves where the depth gain effect of the whole series is taken into account, rather than simply the individual moves. Manoeuvres may be 2-deep, 3-deep, and so on according to the number of moves they contain. Moves are by definition 1-deep manoeuvres. A move may be made in the knowledge that one move eliminates more of a certain type of possibility than another. For example, a bar placed away from the boundary eliminates two possible locations for non-contiguous bars, whereas a bar contiguous with the boundary eliminates only one. This is important, since the location of one bar will often affect where the next can go, and it will turn out that in some processes in the 6 x 6 complex non-edge bars exhaust non-contiguous bars within about fourteen steps, whereas with edge bars it is twenty, and this makes a significant difference to a process. We allow this knowledge within moves, because it can be seen immediately and locally as a consequence of the move, provided the principles are understood.

Both moves and manoeuvres thus have foresight about depth gain, but only manoeuvres have foresight about future moves. A random barring process is one in which barring moves are made independently of each other and without regard for depth gain or any other consequence. We might say then that in describing moves and manoeuvres we are describing the degree to which a process is governed by forethought. At the opposite extreme from the random process, it follows, there will be the process governed by an n-deep manoeuvre, where n is the number of bar locations available, meaning that the whole set of bars is thought out in advance,

Figure 8.11



and each takes into account the known future positions of all others.

Let us now consider different types of barring process. Figure 8.11a–d sets out a barring process of 24 bars, numbered in order of placement in which each move is designed to maximise depth gain. We choose 24 because 25 is the maximum that can be placed without dividing the aggregate into discontinuous zones (that is, in effect, into two buildings), and one less means that one ‘ring’ will remain in the circulation system (that is, one cycle in its graph), so that if there is a process which maximises some property of this ring then we might find out what it is. Bars are numbered in order of placing, and we will now review this ordering.

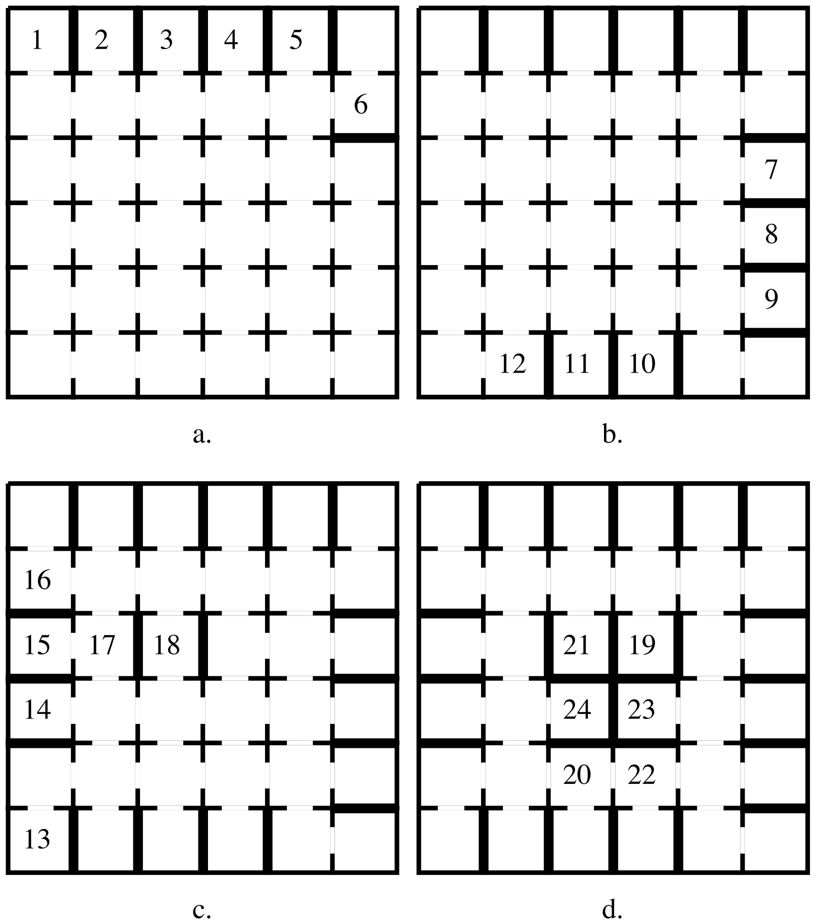
To maximise depth gain, our first bar – bar 1 – must be placed exactly to bisect a line of cells. It does not matter which we select, since the effect of all such bisections will be equivalent. But bar 2 must take into account the location of the first, since depth gain will be maximal only if it is linearly contiguous with it. The same principle governs the location of the bars 3, 4 and 5. After five moves therefore we must have a long central bar reaching to one edge, and we have in fact created the form shown in 8.7g, which is the most depth gain efficient way of using fewest bars to ‘nearly divide’ the aggregate into two. Thus we have arrived at a significant global outcome for the object as a whole, even though we have at each stage only followed a purely local rule. Although individual moves had a certain degree of choice, the configurational outcome as a whole, we can see, was quite deterministic.

Since the next move cannot continue on the central bar line without cutting the aggregate into two, we must look around for the next depth maximising move. We know we must bisect the longest sequence of cells, and if possible our bar must be contiguous with bars already placed. To identify the longest sequence, we must recognise that the barring so far has effectively changed the shape of the complex. We could, for example, cut the complex down the line of the central partition and treat it almost as two complexes. As a result, there is now a longest sequence of cells running around both sides of the central partition which does not form a single line, but it does constitute the longest sequence of shortest available routes in the complex. It is by partitioning this line close to its centre that we will maximise depth gain, that means placing the bar at right angles to the partitioning line at its base in one of the two possible locations. The next bar must then take account of which has been selected, and in fact extend that bar. The next two must repeat the same move on the other side, thus taking us up the ninth bar in the figure. The same principle can then be applied to the next sequence of bars, and in fact all we must do to complete the process is to continue applying the same principle in new situations as they arise from the barring process. By bar 24, the pattern is as shown in the final form in figure 8.11d.

Looking at the final form, we first confirm that once a 25th bar is added no further bar could be added without splitting the aggregate into two. We also note that the configuration of space created by the barring is, excepting the small ring that would be eliminated by bar 25, a single ‘unilinear’ sequence of cells, that is, the form with the maximum possible depth from all points to all others. By maximising

Is architecture an ars combinatoria?

Figure 8.12



depth gain at every stage of the process we arrive, perhaps not surprisingly, at a form which globally maximises depth gain. We also note, that by applying simple rules to the barring process, we have converted a process which theoretically could lead to an astronomical number of possible global forms, to one which leads almost deterministically to a specific form.

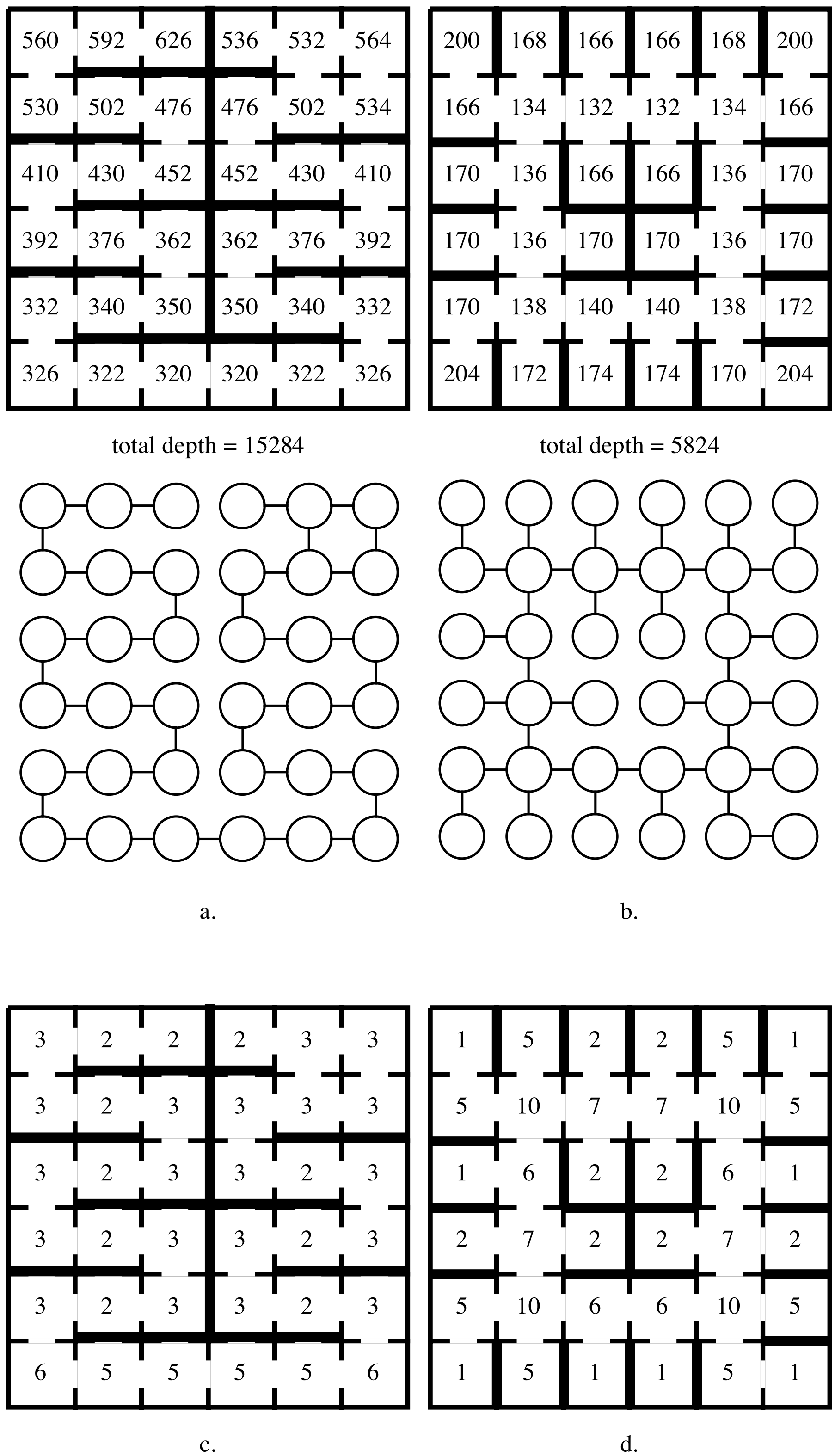
Figure 8.12a–d now illustrates the contrary process in which each move minimises depth gain, again with numbering in the order of the moves. Bar 1 must be at the edge of a line of cells, and to minimise the loss of non-contiguous bar locations it should also be on one of the outermost lines of cells. Once we have bar 1, the following moves to minimise depth gain must continue to bar the already barred line, since this line is now shorter than any other line, and to do so each time as close to the edge of the remaining cell sequence as possible. As before, then, bars 1 – 5 are forced, and lead to a very specific overall pattern. A similar procedure is then forced on other edge lines, obviously omitting bars which would form a right angle with existing bars, since this would split the system into two. Bars 1–16 therefore continue this process until the possibilities are exhausted.

The next move must be non-contiguous and must be as near the edge as possible. Several identical possibilities exist, so we select 17. 18 and 19 must continue to bar the same line, leaving only one of two possible identical further non-contiguous moves. We select 20. Now no more non-contiguous moves are available, so we must select contiguous moves with the least depth gain. The best turns out to be that rebarring the already barred line on which 20 lies has the least depth gain, in spite of the fact that it creates a three-sided enclosure. But the next move cannot create the same pattern to the right, since this will also create a double line block as well as a three-sided enclosure. Barring the open line at 22 has less depth gain than barring the adjacent line to the right, at which point 23 becomes optimal. The final bar must then be on one of five still open lines, the four comprising the 'ring', and the one passing through the centre. Cutting the ring creates much more depth gain than cutting the centre line, because it creates a block in the system that is four cells deep from the boundary. Of the possible locations on the centre line, the central location has less depth gain because the location one to the right creates a two-deep enclosure, which creates more extra depth than the difference between the centre and one-from-centre location.

The depth minimising process has thus given rise to a form which is as striking as the depth maximising process: a ring of open cells accessing outer and inner groups of one-deep cells. We have only to convert the doors in the ring to full width permeabilities to create a fundamental building form: the ring corridor accessing separate 'rooms' on either side. This has happened because the depth gain minimising strategy tends to two kinds of linearity: a linearity in dividing lines of cells up into separate single cells; and a linearity in creating the open cell sequences that provide access to these cells. Aficionados of Ockam's razor will note that both these contrary effects follow from the single rule that bars should always be placed so as to bar the shortest line of cells available as near the edge as possible. This means that once a line has been divided, then it minimises depth gain to divide it again, since, other things being equal, the remainder of an already barred line will always be shorter than an unbarred line. Figures 8.13a and b show typical forms from the two processes, together with depth values for each cell. In fact, the two forms shown in Figures 8.1b and c. The total depth for the near depth maximising process is 15320 while that for the depth minimising process is little more than a third as much at 5824. These differences are all the more remarkable in view of the fact that each form has exactly the same number of partitions. The only difference is the way the partitions are arranged.

But in spite of their differences, each of the forms generated seems in its way quite fundamental. The depth maximising form is close to being a unilinear sequence, that is the form with the maximum possible depth from all cells. The depth minimising form approximates if not a bush, then at least a bush like arrangement built on a ring. We have arrived at these forms by constraining the combinatorial process down certain pathways by some quite simple rules. These have created well defined outcomes through morphological processes which are objective in the sense that although the selection and implementation of rules is a human decision, the local to global morphological effects of these rules,

Figure 8.13


560 592 626 536 532 564
530 502 476 476 502 534
410 430 452 452 430 410
392 376 362 362 376 392
332 340 350 350 340 332
326 322 320 320 322 326
total depth = 15284
200 168 166 166 168 200
166 134 132 132 134 166
170 136 166 166 136 170
170 136 170 170 136 170
170 138 140 140 138 172
204 172 174 174 170 204
total depth = 5824
a.
b.
3 2 2 2 3 3
3 2 3 3 3 3
3 2 3 3 2 3
3 2 3 3 2 3
3 2 3 3 2 3
6 5 5 5 5 6
1 5 2 2 5 1
5 10 7 7 10 5
1 6 2 2 6 1
2 7 2 2 7 2
5 10 6 6 10 5
1 5 1 1 5 1
c.
d.

whether for the individual move in the process or the accumulative result, is quite independent of human decision. The eventual global pattern of space 'emerges' from the localised step-by-step process. At the same time, processes whose rules are similar 'converge' on particular global types which may vary in detail but at least some of whose most general properties will be invariant – the tendency to form long sequences with few branches, the tendency to generate one-deep dead end spaces, the tendency to form smaller or larger rings and so on.

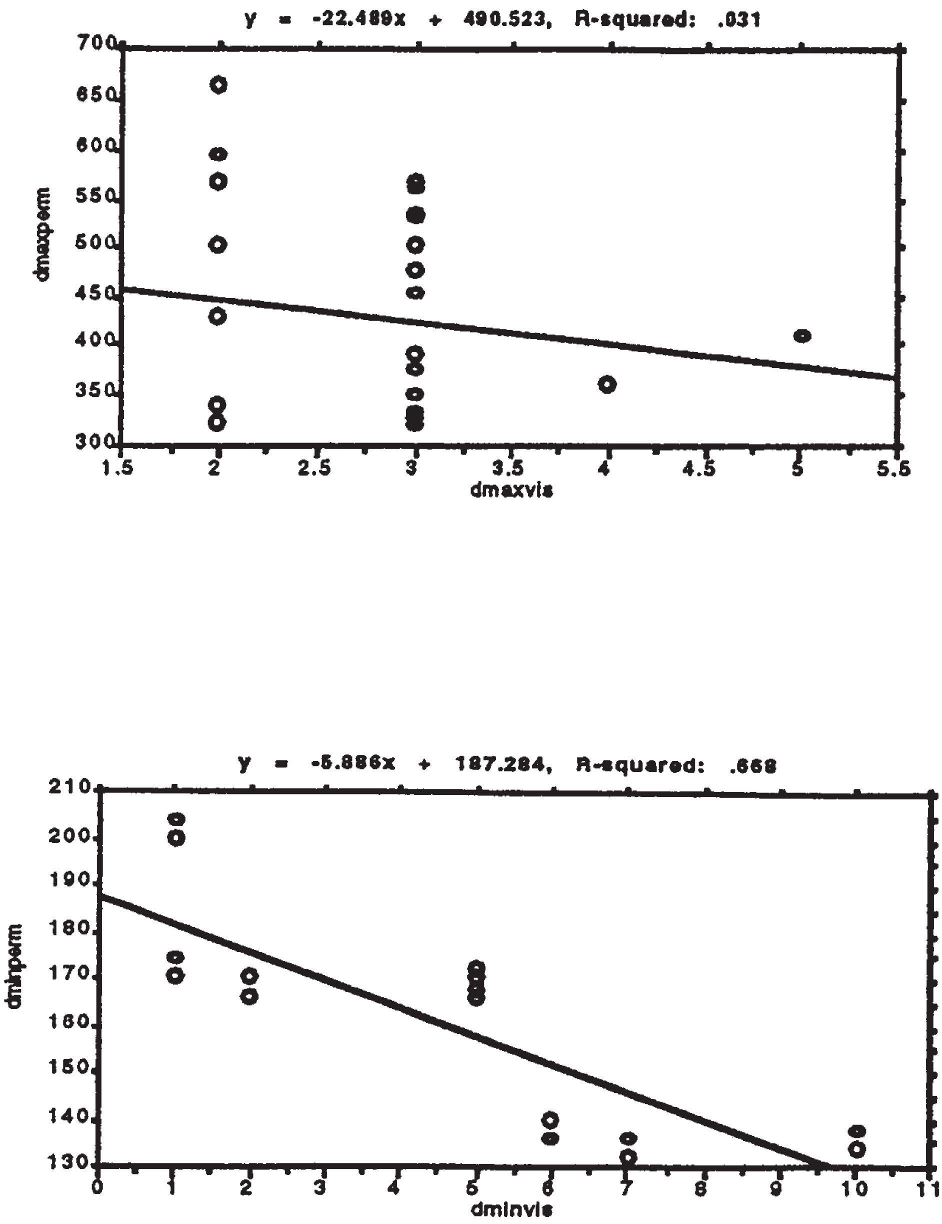
This combination of emergence and convergence is immensely suggestive. It appears to offer a natural solution to the apparent paradox we noted at the start of this chapter: that in spite of the vastness of the combinatorial field, intuition suggested relatively few ways of designing space. We may now reformulate this paradox as a tentative conclusion: consistently applied and simple rules arising from what is and is not an intelligible and functionally useful spatial move create well-defined pathways through the combinatorial field which converge on certain well-defined global spatial types. These laws of 'emergence-convergence' seem to be the source of structure in the field of architectural possibility. What then are these laws about? I propose they are about what I called 'generic function', that is properties of spatial arrangements which all, or at least most, 'well-formed' buildings and built environments have in common, because they arise not from specific functional requirement, that is, specific forms of occupation and specific patterns of movement but from what makes it possible for a complex to support any complex of occupation or any pattern of movement.

The theory of generic function: intelligibility and functionality

The first aspect of generic function reflects the property of 'intelligibility' which Steadman suggests might be one of the critical factors restricting architectural possibility. In Chapter 4 we suggested that the intelligibility of a form can be measured by analysing the relation between how a complex can be seen from its parts and what it is like in an overall pattern, that is, as a distribution of integration. This was expressed by a scattergram showing the degree of correlation between the connectivity of a line, which is a local property of the line and can be seen from the line, and integration, which is a global property relating the line to the system as a whole and which cannot therefore be seen from the line. How might this concept relate to the construction of spatial patterns by physical moves? Visibility is in fact interesting since it behaves in a similar way to depth under partitioning. For linear cell sequences the effect of bars on visibility exactly mirrors depth gain, though in a reverse direction: visibility lost from a bar is exactly half the depth gain from the same bar, and as the bar moves from edge to centre the total visibility along the line decreases, while at the same time the visibility value of cells along that line become more homogeneous, eventually becoming the same with a central bar.

In our two complexes then, let us define visibility very simply as the number of cells that can be seen from the centre of each cell. These visibility values are set out for our two depth maximising and minimising complexes in figures 8.13c and d.

Figure 8.14


y = -22.489x + 490.523, R-squared: .031
dmaxperm
dmaxvis
y = -5.886x + 187.284, R-squared: .668
dminperm
dminvis

Is architecture an ars combinatoria?

These visibility values and their mean index the visual connectivity of the complex. We may also express these by drawing an axial map of the fewest lines that pass through all the cells. We can see how many cells each line passes through, and how this differs from one complex to another. We can if we wish express this in a summary way by working out the ratio of the means depths for each cell and the mean visibility of each cell. For the depth minimising form, the mean depth from cells is 5.3, and the mean visibility is 3.9. We might call this a .74 visibility to depth ratio. In the depth maximising form, the mean depth is 11.9 while the mean visibility is 2.8, a visibility ratio of .24, about a third that of the depth minimising form. This seems to agree quite well with intuition.

This shows how the visibility and depth properties of the complex relate to each other. However, we may learn more by correlating the permeable depth figures for cells with their visibility figures and expressing the relation in a scattergram. The better the values correlate, the more we can say that what you can see from the constituent cells of the system is a good guide to the global pattern of depth in the complex which cannot be seen from a cell, but which must be learnt. The correlation thus expresses the intelligibility of the complex. Figures 8.14a and b are the scatters and correlation coefficients for our two cases, showing that the depth minimising form is far more intelligible than the depth maximising form. This formally confirms our intuition that the depth maximising form is hard to understand, in spite of being a single sequence, because the sequence is coiled up and the information available from its constituent cells is too poor and undifferentiated to give much guidance about the structure of the complex as a whole from its parts. The opposite is the case in the depth minimising complex. On reflection, we can see that this will always tend to be the case with depth maximising processes since the partitioning moves that maximise depth are also those which also maximally restrict visibility.

There are therefore, as Steadman suggests, fundamental reasons to do with the nature of human cognition and the nature of spatial complexes which will bias the selection of spatial forms away from depth maximising processes and in the direction of depth minimising processes. Through this objective – in the sense that we have measured as a property of objects rather than as a property of minds – property of intelligibility then we can see one aspect of generic function structuring the pathways from combinatorial possibility to the architecturally real.

There are, however, further reasons why depth minimising forms will be preferred to depth maximising forms which have to do with functionality. Functionality we define as the ability of a complex to accommodate functions in general, and therefore potentially a range of different functions, rather than any specific function. Intuitively, deep tree-like forms such as the depth maximising form seem functionally inflexible and unsuited to most types of functional pattern while the depth minimising form seems to be flexible and suited to a rather large number of possible functions. Can this be formalised?

Is architecture an ars combinatoria?

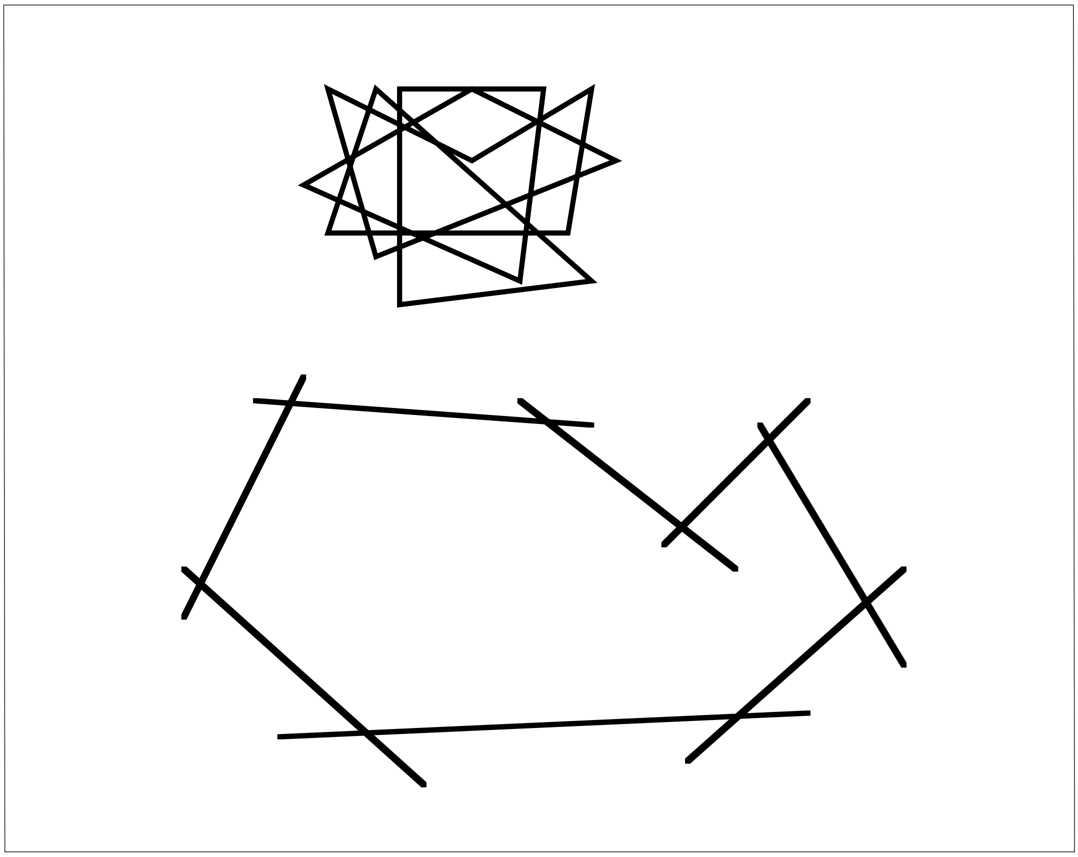


Figure 8.15 (above)
Locally convex movement; when small movements intersect and form a local convex region.

Figure 8.15 (below)
Globally linear movement; when large scale movement forms strings or rings of lines.

It is useful to begin by considering in as generic a way as possible the types of human behaviour that occur in buildings. We may do this best by considering not the purpose or meaning of an activity but simply its physical and spatial manifestation, that is, what can actually be observed about human activity by, say, an extra-terrestrial who had no idea what was going on and could only record observations. Generically, such an observer would conclude, two kinds of thing happen in space: occupation and movement. Occupation means the use of space for activities which are at least partly and often largely static, such as conversing, meeting, reading, eating or sleeping, or at most involve movement which, when traced over a period, remains localised within the occupied space, such as cooking or working at a laboratory bench, as shown in figure 8.15.

Movement we can define not as the small local movements that may be associated with some forms of occupation, and therefore to be seen as aspects of occupation, but movement between spaces of occupation, or movement in and out of a complex of such spaces. Movement is primarily about the relations between spaces rather than the spaces themselves, in contrast to occupation which makes use of the spaces themselves. We can see this as a scale difference. Occupation uses the local properties of specific spaces, movement the more global properties of the pattern of spaces.

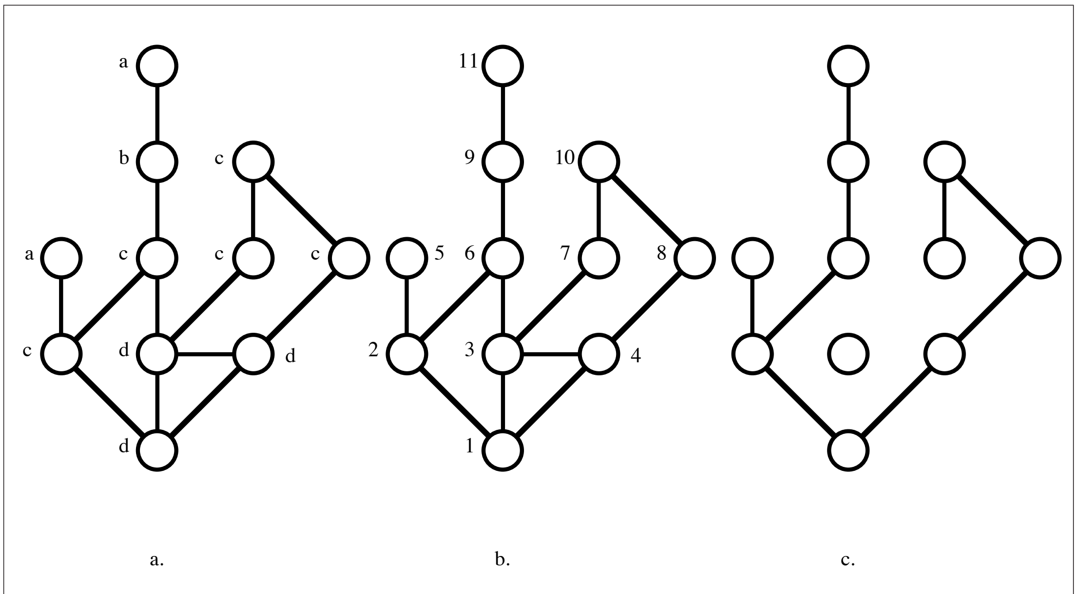
There is also a difference between occupation and movement in the spatial form each takes. Because spatial occupation is static, or involves only localised

Is architecture an ars combinatoria?

movement, the requirement that it places on space is broadly speaking convex, even when this involves localised movement within the space. In particular, any activity that involves the interaction or co-presence of several people is by definition likely to be convex, since it is only in a convex space that each person can be aware of all the others. Movement, on the other hand, is essentially linear, and the requirement that it places on space is consequently linear, at least when seen locally in its relation to occupation. There must be clear and relatively unimpeded lines through spaces if movement is to be intelligible and efficient.

Occupation and movement then make requirements of space that are fundamentally different from each other in that one is convex and the other linear. Because this is so there is an extra difficulty in combining occupation and movement in the same space. There will always, of course, be practical or cultural reasons why different forms of occupation cannot be put in the same space – interference, scaling of spaces, privacy needs, and so on – in spite of the fact that each is convex and in principle could be spatially juxtaposed to others. But to assemble movement and forms of occupation in the same space is in principle more difficult because, over and above functional interference, occupation and movement have fundamentally

Figure 8.16



different spatial shapes. The interference effect from occupation to occupation and from movement to movement will be of a different kind to that from occupation to movement because the spatial requirements are more difficult to reconcile.

Because this is so, it is common to find that the relation between movement and occupation in spatial complexes is often one of adjacency rather than overlap, whether this occurs in spaces which are fully open (as for example when we have both lines of movement and static occupation in a public square), or fully closed, as when we have rooms adjacent to corridors, or one is open and the other closed, as when houses align streets. In each case, the linearity required for movement is achieved by designing movement to occur in spaces which pass immediately by rather than through occupation spaces.

Now let us consider the types of space that are available to meet the requirements of occupation and movement. First we must consider the most basic topological properties as embodied in the graph of a complex, since even at this level topologically different types of space have quite different potentials for occupation and movement. Let us first consider, a familiar graph, as shown in 8.16a, b and c.

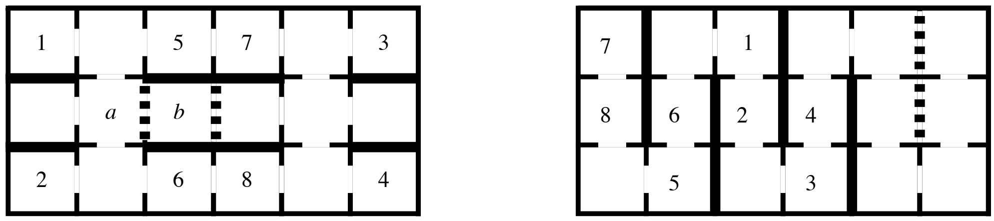
In this graph, as in others, the spaces that make up the graph can be divided into four topological types. First, there are spaces with a single link. These are by definition dead-end spaces through which no movement is possible to other spaces. Such spaces have movement only to and from themselves, and are therefore in their topological nature occupation-only spaces. Examples are marked *'a'* in figure 8.16a. The link from one-connected spaces to the rest of the graph is necessarily a cut link, meaning that its elimination must split the graph into two, in this case the space whose link has been cut and the rest of the graph. Because the cut link only serves a single space, the effect of cutting makes little difference to the remainder of the complex beyond minor reductions in the depth of the rest of the complex following the elimination of a space.

Second, there are spaces with more than one link but which form part of a connected sub-complex in which the number of links is one less than the number of spaces, that is, a complex which has the topological form of a tree. Such spaces cannot in themselves be dead end spaces, but must be on the way to (and back from) at least one dead end space. All links to spaces in such complexes, regardless of the number of links to each space, are also 'cut links' in that the elimination of any one link has the effect of splitting one or more spaces from the rest of the complex. Such spaces are marked *'b'* in figure 8.16a. A consequence of the definition is that there is in any such sub-complex (or complex) exactly one route from each space to every other space, however large the sub-complex and however it is defined. This implies that movement through each constituent space will only be to or from a specific space or series of spaces. This in turn implies that movement from origins to destinations which necessarily pass through a b-type space must also return to the origin through the same space.

Third, there are spaces with more than one link which form part of a connected sub-complex which contains neither type a nor type b spaces, and in which there are exactly the same number of links as spaces. Such spaces are marked c in figure 8.16a. The definition means that c-type spaces must lie on a single ring (though not all spaces on the ring will be c-type) so that cutting a link to a c-type space will automatically reduce the ring to one or more trees. Movement from a c-type space through a neighbour need not return through the same neighbour but must return through exactly one other neighbour.

Finally there are spaces with more than two links and which form part of complexes which contain neither a- nor b-type spaces, and which therefore must contain at least two rings which have at least one space in common. Such spaces

Figure 8.17



a = 1102

b = 1082

a. b.

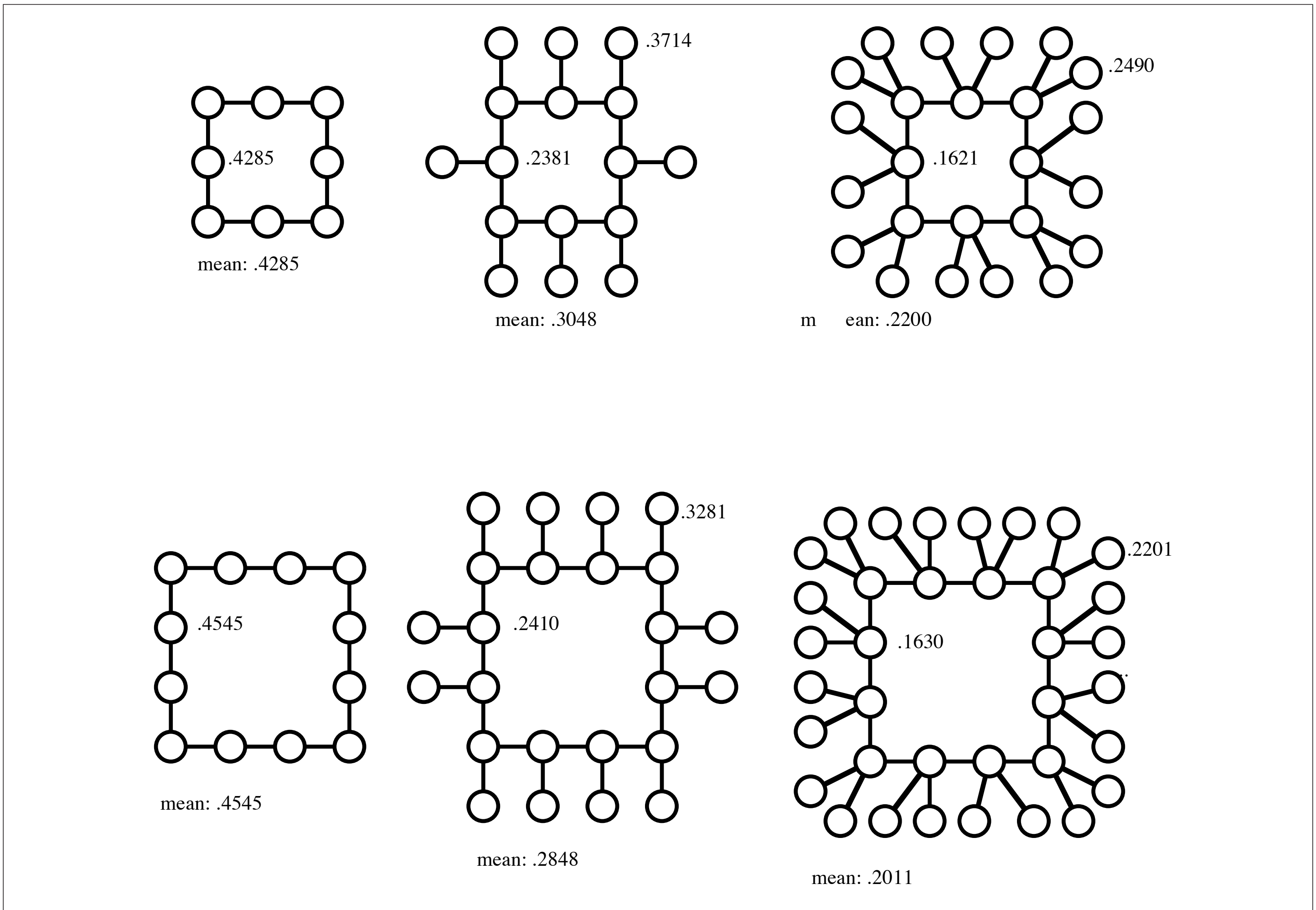
must lie on more than one ring, and are labelled 'd' in figure 8.16a. Movement from d-type spaces through a neighbour has the choice of returning by way of more than one other neighbour.

We may also define subcomplexes of the a-, b-, c- or d-type as the space of that type plus all the spaces by reference to which it is defined as a space of that type, even though some of those spaces may belong also to other subcomplexes. (In other words, a subcomplex of a given type is a complex containing at least one space of that type.) Looking at numbered spaces in figure 8.16b, we can then say that spaces 5 and 11 are a-type spaces, and that the sub-complex formed by spaces 2 and 5 and that formed by 9 and 11 can be thought of as a-type subcomplexes. Space 9 is a b-type space, and that the subcomplex formed by spaces 6, 9 and 11 can be seen as a b-type sub-complex. Spaces 2, 6, 7, 8 and 10 are c-type spaces and each may be seen as forming part of a local ring, or c-type complex: thus 2 and 6 are part of the c-type subcomplex formed by spaces 1, 2, 6 and 3, and 7, 8 and 10 are part of the c-type complex formed by spaces 3, 7, 10, 8 and 4. Space 3 and 4 are d-type spaces and are part of the d-type subcomplex formed by spaces 1, 2, 3, 4, 6, 7, 8 and 10. Spaces are, in effect, unambiguously defined by their place in a complex, but this does not mean that spaces that contribute to that definition do not form part of other complexes. For example, an a-space may be part of a b-complex, or a c-space may be part of a d-complex without in either case compromising its unique identity as an a- or c- type space.

There are simple and fundamental relationships between these elementary topologies and the depth minimising and maximising processes. A depth minimising process will in its nature tend first to leave long lines of spaces unimpeded and to preserve their connection to other long lines, and second to coil contiguous bars up into small, one-deep 'rooms'. This is illustrated in figure 8.17a where the first eight bars cut the shortest lines, to create rooms at either end and potential rooms in the centre. The dotted bars marked 'a' and 'b' represent two possible choices at this point, and the figure on the right side shows the total depth in the system after each. The analysis shows that the two one-deep rooms add far less depth than one two-deep complex, in effect because the two-deep complex is created by five contiguous bars, whereas the one-deep spaces are each created by three contiguous bars. The depth minimising process thus tends to create a-type spaces linked by global c- and d-type complexes, as was the case in the 6 x 6 example in figure 8.13b. In contrast, the depth maximising process, as shown in figure 8.17b for example, will by contiguously barring the longest available lines, create b-type spaces and therefore sequences rather than a-type spaces, and localise c- and d-complexes at the earliest possible stage of generation, and with a configuration in which there are few a-type spaces, and these at the end of long sequences, with any rings in the system highly localised.

In other words depth minimising processes will tend locally to a-type complexes and globally to d-type complexes (in figure 8.13b it is only the final 24th bar that reduces a strong global d-complex to a global c-complex), while depth maximising processes will tend globally to b-type complexes and locally to small residual c-type complexes. This is instructive because it tells us how these elementary configurations are related to the product of the functionally critical property of integration in spatial complexes. Essentially, a- and d-type spaces create integration, while b- and c-type spaces create segregation. In other words, segregation in a complex is created almost entirely by the sequencing of spaces. Since this is not obvious, it is worth illustrating. In figure 8.18 for example, in the left column, we increase the size of the ring from 8 to 12 spaces and the i-value increases (i.e. becomes less integrated) from .4285 for the 8-ring to .4545 for the

Figure 8.18



12-ring. In the second column, we add a single a-type space to each c-type space. Both complexes become on average more integrated, but the 12-ring complex below becomes relatively more integrated at .2848 than the 8-ring complex above at .3048. In fact, the ring spaces in the 12-ring complex are slightly less integrated at .2410 than those of the 8-ring complex at .2381, but the a-type space of the 12-ring complex are markedly more integrated at .3281 than those of the 8-ring complex at .3714. In the right column, we link two a-spaces to each c-space and the pattern becomes even more marked. The 12-ring complex is now more integrated at .2011 than the 8-ring complex at .2200, with ring spaces at .1630 compared to .1621, but a-spaces at .2201 compared to .2490.

We now have a more or less complete account of the relation between generative processes, the creation of different types of local and global space complexes, and the construction of patterns of integration. We can now formulate the question at the centre of our argument: what are the implications of these spatial variations for occupation and movement, that is, for the generic functioning of spatial complexes? In exploring this, we should bear in mind one of the major findings of the research reported in Chapters 5 to 8: that the more movement in a complex is from all parts to all other parts, then the more the pattern of movement in a complex will tend to follow the pattern of integration.

First we must note that each of the types of space we have identified, and the type of complex it characterises, has generically different implications for space occupation and movement. As we have already indicated, a-type spaces do not have through movement at all and therefore do raise the issue of relating occupation to movement (other than movement to and from the space itself). b-type spaces raise the possibility of through movement but also control it strongly, both because each route through a b-type space is unique and also because return movement must pass through the same space. c-type spaces also raise the possibility of through movement while also constraining it to specific sequences of spaces, though without the same requirement for the return journey. d-type space permits movement, but with much less built-in control because there is always choice of routes in both directions.

It is clear then that b-type and to a lesser extent c-type spaces have a much more determinative relation to movement than either a-type or d-type spaces. While the a-type does not allow for through movement, and the d-type allows choice of movement, the b-type and the c-type permit but at the same time constrain it by requiring it to pass through specific sequences of spaces. The b-type is the most constraining. For any trip from an origin to a destination, every b-space offers exactly one way in and one way out of each space and every trip in a b-complex must pass both ways through exactly the same sequence of spaces. A similar, though weaker, effect is found for c-spaces and c-complexes, because although at the level of the ring as a whole there will be a choice of one direction or another, trips once begun must use a single sequence of spaces, and the trip therefore resembles a b-trip, though without the requirement that the return journey repeat the same sequence in reverse. This effect arises from the simple fact that b- and c-type spaces are from the point of view of any trip that passes through them, effectively two-connected, and two is the smallest number that allows entry to a space in one direction and egress in another. It is this essential two-connectedness from the point of view of trips, that gives b and c-spaces their distinctive characteristic of both permitting and constraining movement.

Now this means that b- and c-type spaces raise issues for the relation between occupation and movement which are not raised either by one-connected or more than two-connected space, in that they require the resolution of the relation between occupation and through movement within each convex space. This has a powerful effect on the usability of spaces and space complexes of this kind. In general, it can only occur where the sequencing of spaces reflects a parallel functional sequencing of occupation zones, and movement is, as it were, internalised into the functional complex and made part of its operation.

For example many types of religious building use exactly this spatial property to create a sequence of spaces from the least to the most sacred, each space having different occupational characteristics. More commonly, we find the phenomenon of the ante-room, for example where a senior person in an organisation places a subordinate in a space which controls access to the office. In domestic space, such interdependencies are quite common. Indeed, the domestic dwelling may often be characterised as a pattern of such interdependencies. Figure 8.16, for example, has a maximally simple b-complex (spaces 6, 9 and 11) associated with male working activity and a near maximally simple c-complex (spaces 3, 7, 10, 8 and 4) associated with female working activity, as well as a maximally simple a-complex (spaces 2 and 5) associated with formal reception and a dominant d-type space (space 3 – the salle commune) in which all everyday living functions, including informal reception, are concentrated and which holds the whole complex together. It is notable that if this space (space 3) is removed from the complex, as in figure 8.16c, the whole complex is reduced to a single sequence with a single one-deep branch.[9]

In general we can say that the sequencing of spaces normally occurs when (and perhaps only when) there are culturally or practically sanctioned functional interdependencies between occupation zones which require movement to be an essential aspect of these interdependencies and therefore to be internalised into a local functional complex of spaces.[10] Such interdependencies are comparatively rare and, because they are so, where they do occur they tend to be highly localised. There are simple combinatorial reasons for this. If interdependency requiring internalisation of movement into a functional complex is unusual for pairs of occupation types, it is even more unusual for triples, even more for quadruples, and so on. This is why it tends to remain localised.

It follows that whereas in small buildings, such functionally interdependent complexes can form a significant proportion of the complex, or even the whole complex, as buildings grow large and acquire more and more occupation spaces, those that have the necessary interdependencies that require spatial sequencing will become a diminishing proportion of the whole. As buildings grow therefore more and more of the movement will not be of the kind which is internal to the functioning of a local subcomplex but will occur between subcomplexes which are functionally much more independent of each other.

This means that movement will be less 'programmed', that is, a necessary aspect of interdependent functions, and more contingent, or 'unprogrammed'.[11] It follows that the pattern of movement will follow from two things: first from the way in which the various occupation spaces are disposed in the spatial complex, coupled to the degree to which each acts as an origin and a destination for movement between occupation spaces; second, from how this disposition relates to the spatial configuration of the complex itself. The more movement occurs more or less randomly from all locations (or even all parts of the complex) to all others, then the more it will approximate the conditions that give rise to 'natural movement', that is movement through spaces generated by the configuration of space itself, and the more movement will then follow the pattern of integration of the building. The more this occurs, the more movement will be functionally neutralised, that is, it will not be an intrinsic aspect of local functional complexes determined by the functional programme of the building but as a global emergent phenomenon generated by the structure of space in the building and the disposition of occupation spaces within it.

Neutralised movement will then tend to follow the configurational topologies that generate the pattern of integration in a building. a-space will have no movement other than that starting and finishing in them; b-space will have movement only to the spaces to which they control both access and egress; c-spaces will have movement to spaces to which they control either access or egress; while d-spaces will be natural attractors of movement. It follows that just as a-spaces are the most suited for occupation because they are least suited for movement, so d-spaces are the least suited for occupation, because they are the most suited to movement, especially where this movement is from all locations to all other locations in the complex.

It follows that a growing spatial complex will need a decreasing proportion of b- and c-complexes since these will only be needed for local functionally interdependent groups of occupation spaces, and a growing proportion of a-type and d-type complexes. In such complexes there will be a natural specialisation of spaces into a-complexes for occupation and d-complexes for movement, and therefore an equally natural tendency towards the adjacency relation for occupation and movement.

As we have seen, it is exactly such complexes that are generated by depth minimising processes. Such complexes also have other advantages. First, because the mix of a-type and d-type complexes is in its nature the most integrated, then journeys from all spaces to all others will be on average topologically (and in fact metrically) shorter than for any other type of complex. Second, such complexes maximise the number of a-spaces for occupation while minimising the number of spaces in the d-complex for movement, thus making the relation of occupation and movement as effort-efficient as possible. Third, the more this is the case, the more movement from specific origins to specific destinations in the complex will overlap and create a global pattern of co-presence and co-awareness of those who are not brought together in the local functional subcomplexes of the building. In other words, the movement pattern brings together in space what the occupational requirement of the complex divides. This reflects the basic fact that whereas the overlap of occupation type in the same space is likely to cause interference from one to the other, the overlap of movement in situations where movement is functionally neutralised creates an emergent form of spatial use – co-presence through movement – which is essentially all of the same type. Overlap is therefore not likely to be read as interference. On the contrary, it is likely to be read as a benefit.

It is then in the nature of things that spatial complexes of this type will tend to become dominant as buildings grow in scale and occupational complexity. This type of configuration arises from generic function, that is, from the fact of occupation and the fact of movement, prior to any consideration of the specific functions to be accommodated in the building. We only need to add the larger open spaces and longer linear spaces in the d-complex in accordance with the principles we have established to optimise the relation between occupation and movement in the complexes.

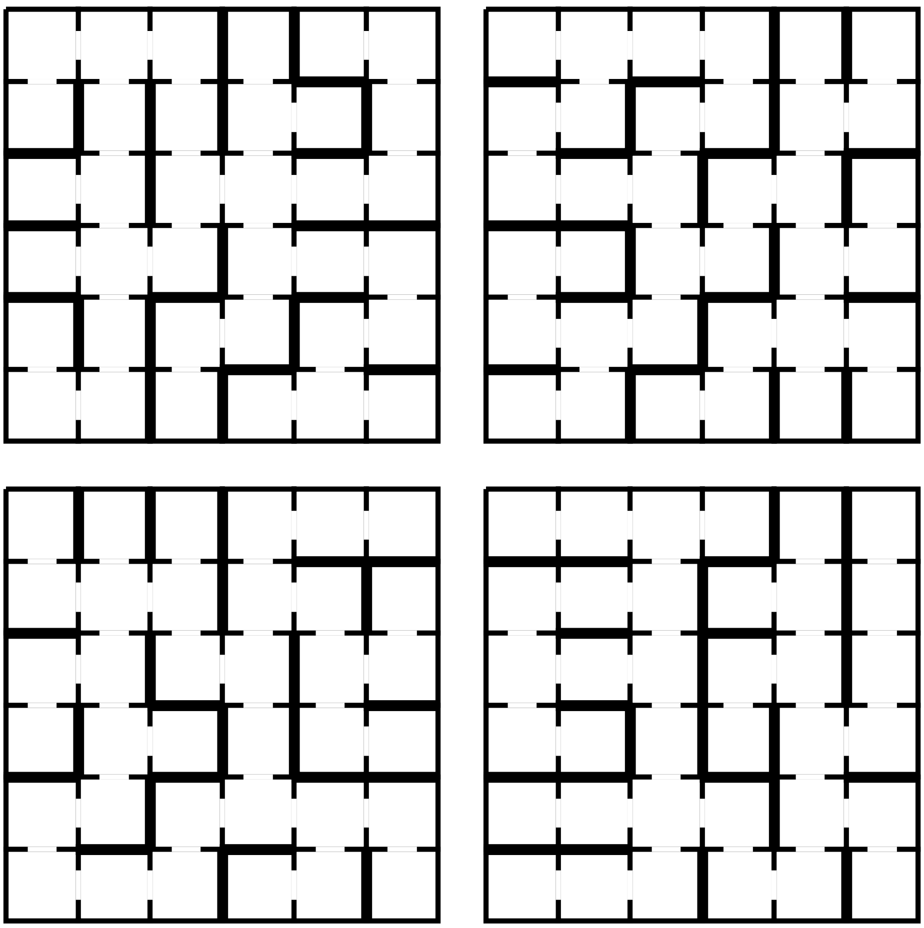
So, is architecture an ars combinatoria?

We have now answered the question asked at the beginning of the chapter, and embodied in the two prefatory quotes. No theory of architecture as an *ars combinatoria* of elements and relations is useful because, as with language, it is how combinatorial possibility is restricted that gives rise both to the 'structure of the language' and to the 'elements' of which the language is composed. The vast majority of combinatorial possibilities are as irrelevant to that language as random sequences of words are to natural language. The structure of the language, which eliminates most possibilities, arises not from basic rules for combining basic elements, but from local to global laws from physical moves to spatial configuration, which give rise at one level to the local stabilities we call elements and at another to the higher order patterns that characterise the general spatial forms of buildings.

The effects of understanding how restrictions on combinatorial possibility create the 'language of space' are two. First, we see that there are not in any useful sense basic elements. Elements arise from local spatial strategies that realise – and must then be taken as intending to realise – particular local to global spatial ends. All are describable as spatial phenomena emergent from the consistent application of rules governing either the completion or removal of a single type of fundamental spatio-physical element: the permeable partition. It is the record of this consistent application that we see when we name a local configuration as a certain kind of element. If we randomly partition a complex, as in the four examples in figure 8.19, we do not find such consistencies, and we are not therefore inclined to identify elements. We should properly see 'elements' as 'genotypes', that is, systems of informational abstractions governing objects whose phenotypes are endlessly varied. It is only in this way that we can reconcile the idea of a well-formed 'element' with the fact that such elements arise from and are given by configurational relations, not only those which generate their intrinsic form, but also those which define their embedding in the system as a whole. In one sense we might say

Is architecture an ars combinatoria?

Figure 8.19



that we have reduced the apparent fundamental elements of spatial complexes to something more elementary: a small family of local physical moves which by following different rules produce spatial effects in the complex. But in a more important sense, we have dissolved the element into two sets of configurational laws: the laws that generate the element itself, and those that generate the impact of the element on the complex as a whole.

Second, we see that it is not useful to think of global patterns as arising simply from relations among elements. In a spatial configuration, every local move has its own configurational effect, and it is the natural laws that govern these local to global effects that govern global configuration. It follows that it is knowledge of these laws that we require for a theory of space, not knowledge of combinatorial possibility. It is these laws that give rise to both the local configurational types we are tempted to call elements and to the global configurational patterns that commonly characterise buildings as a whole. We can thus solve the apparent paradox of vast combinatorial possibility and a few basic pattern types. It is the natural local to global laws restricting possibility that lead space to converge on the pattern types that we find.

The precise form of these laws governing the relation between possible spatial configuration and generic function lies in the fact that individual, localised design moves – say making a partition, or eliminating a doorway – have global configurational effects, that is, effects on the overall pattern of space. These global pattern effects of local moves are systematic, so that different types of move, carried out consistently, will give rise to very different configurational effects. These local to global laws are independent of human volition, and as such must be regarded as more akin to natural laws than contingent matters of human existence. This does not imply that the relationship of human beings to space is governed by natural laws, but it does mean that the passage from the possible to the actual passes through – and has historically passed through – natural laws which mediate the relationship of human beings to space. The built forms that actually exist, and have existed, are not, as they are often taken to be, simply subsets of the possible, but variable expressions of the laws that govern the transition from the possible to the real. These laws, and their relation to generic function, are therefore the true constraints on spatial possibility in architecture and urban design, and a theory of space must be an account of these laws.

Does this mean we should abandon combinatorics altogether? We should not. Combinatoric possibility is the framework within which architectural actuality exists, and the proper form of a theory is one that describes how possibility becomes actuality. We are now in a position to suggest the general framework for such a theory. The huge number of possible spatial arrangements, we suggest, pass through a series of three filters before they become real buildings. The filters operate at different levels, but all have to do with the human purposes for which we make buildings; that is, these filters are functional filters of possible forms.
The first filter is the most general: that of generic function, as we have described above. This governs the properties which all spatial arrangements must have in order to be usable and intelligible to human beings at all, that is, in order for human beings to be able to occupy space, to move about between spaces and to find buildings intelligible. The second filter is the filter of cultural intent. This refers to the way in which buildings tend to form culturally defined types so that buildings which perform the same culturally defined function in a specific time and space tend to have at least some common spatial properties. We may call this filter that of the cultural genotype. The third filter is the level of the specific building, where those aspects which are not specified by the cultural genotype can vary either in a structured or random way, giving rise to individual differences in buildings. These three functional filters are not independent of each other, but work in succession. For example, all level-two cultural genotypes work within the limits set by the generic function filter of level-one. Similarly, level-three filters work within the constraints set at level-two.

There is, however, a further reason why we should not abandon combinatorics. Although we have shown in this chapter that the combinatorial study of formal and spatial possibility in architecture cannot in itself lead to the theory of architectural possibility, this does not end the matter. Although the theoretical space of buildings is only a part of the theoretical space of spatial combinatorics, it nevertheless is a part of that field, and as such it must obeys its laws. If this is the case, then we find that having eliminated combinatorics as a theory of architecture, we must re-admit it as meta-theory.

Let us argue from a precise example. In Chapter 2, we discussed a thought experiment called the 'Ehrenfest game' as a model for the concept of entropy. In this experiment, 100 numbered balls placed in one jar eventually get more or less evenly distributed between two jars if we randomly select a number and transfer the corresponding ball from whichever jar it is in to the other. This happens because the half and half state is the most probable state because there are far more microstates, that is, actual distributions of the numbered balls, corresponding to the half and half macrostate (that is the actual number of balls in each) than to macrostates in which the balls are unevenly distributed. The shifting probabilities of this process give an insight into the formal nature of 'entropy'.

Now the point of the 'Ehrenfest game' is that it is a useful analogue for the physical notion of 'entropy', as found for example in mixing gases. It is relevant to our argument because we can use the Ehrenfest model to explore a random partitioning process, and in doing so learn important lessons about partitioning in general. All we need do is set up a process for randomly partitioning our spatial complex by numbering our 60 partitions in the 6×6 complex and setting up the random selector to select a number between 1 and 60. We then spin the pointer to select numbers in succession, and each time a number is selected go to the partition with that number and change its state; that is, open a doorway in a partition without one, and close it off if it has one. What happens? Intuition says that the process will eventually settle down to a state in which about half the partitions have doorways and half do not, and that this is therefore the most probable state. We already know that this is the state where there are the maximum possible number of different arrangements.

We may show this, and understand its relevance, by thinking through carefully what will happen in our random process. The first time a number is selected, the probability of opening a doorway rather than closing one is 60/60, or 1, meaning certainty. The second time, there is a 1/60 chance of closing the same door we have just opened (a .0167 probability) and a 59/60 chance of opening another (a .9833 probability). The third time, there is a 2/58 chance of closing one of the doors we have just opened (or a .0345 probability), and a 58/60 chance of opening another (a .9667 probability). Evidently as we progress, the chances of closing a door rather than opening another begin to approach each other until when we have 30 doorways open and 30 partitions closed, the chances are exactly equal. Opening and closing doors are therefore 'equiprobable'.

In other words, we have the same type of combinatorics for a partitioning process as we do for an Ehrenfest game, and therefore for the concept of entropy. This conclusion has clear architectural implications. For example, it explains that, as we have already noted, there are far more partitioning states for about half the number of possible partitions than there are for smaller or larger numbers. There is then a greater range of states for partitioning close to the maximum for a single complex (as in the depth maximising and depth minimising examples) and it is also in this region that small changes to a partitioning have the maximum effect on the distribution of integration, as for example moving a single partition to cut a large ring. There are a whole family of such and similar questions which arise from the basic combinatorics of space, even though buildings occupy only a small part of the combinatorial range.

The laws of spatial combinatorics are not therefore the spatial theory of architecture but they do govern it and constitute the meta-structure within which the theoretical space of real architectural possibility exists. Spatial combinatorics is therefore the meta-theory of architectural space, not its theory. The relationship is exactly analogous to that between the mathematics of ‘information theory’ and the science of linguistics. The mathematical theory of communication is not itself the theory of language, but it is the meta-theory for the theory of language, because it is the framework of general laws within which linguistic laws come into existence. As with language, mathematical laws of combinatorics are everywhere present in architectural possibility because they are the framework for that system of possibility. They need therefore to be understood as a pervasive, containing framework for the theory of architectural space.

In the next chapter we will see that there is a much more pervasive sense in which combinatorics is the meta-theory of architectural possibility, that is, when we come to study not the discrete sets of possibilities which we have considered so far, but when we look at aggregative processes of the kinds that prevail in urban systems of all kinds, and in building complexes as they become large. Here we will see that, as discussed briefly in Chapter 8, combinatorial probability actually plays a constructive role in architectural morphogenesis.

Notes

1 W. R. Lethaby, *Architecture*, Home University Library, London 1912.
2 M. Hellick, *Varieties of Human Habitation*; 1970
3 E. Trigueiro, *‘Change and Continuity in Domestic Space Design: a comparative study of houses in 19th and early 20th century houses in Britain and Brazil’*, PhD thesis, ucl 1994
4 As reviewed in P. Steadman, *Architectural Morphology*, Pion 1983, Chapter 8 and Appendix.
5 P. Steadman, p 171.
6 Trigueiro, ‘Change and continuity’.
7 These proportions are estimated from the results yielded by the ‘second normalisation’ to large numbers of cases. It must be stressed that they are at this stage only tentative approximations. The general point, however, seems secure.
8 An exactly analogous conclusion about the nature of ‘elements’ in language is reached by de Saussure in *Course in General Linguistics*, McGraw Hill, 1966 (originally in French, 1915). For example: ‘Language does not offer itself as a set

of pre-delimited signs to be studied according to their meaning and arrangement', p.104; 'We are tempted to think so if we start from the notion that the units to be isolated are words...the concrete unit must be sought not in the word, but elsewhere', p.105; and 'Language, in a manner of speaking, is a type of algebra consisting solely of complex terms...language is a form not a substance...all our incorrect ways of naming things that pertain to language stem from the involuntary supposition that the linguistic phenomenon must have substance', p.122.

9 In other words, each kind of occupation is characterised by a distinctive local configuration, dependent for their integration into a single complex on the spatio-functionally central salle commune. It is the fact of being an assemblage of different local sub-complexes into a single configuration that makes the dwelling distinctive as a building type. The dwelling is not, as it is often taken to be, the simplest building. On the contrary, seen as an intricate pattern of functional interdependencies mapped into space, it may well be the most complex.

10 In buildings where the organisation of a specific pattern of movement is a dominant functional requirement we can expect space to be dominated by sequencing. For example, galleries and exhibition complexes, which are designed explicitly to move people through the complex so that all spaces can be traversed without too much repetition, normally have a high proportion of c-type sequenced spaces, giving their justified graphs the distinctive form of a number of deep, intersecting rings. This is not, however, a clear case. If we examine the functional microstructure of gallery spaces we find that the lines of global movement pass through the sequenced space in such a way as to leave the viewing zones free for only local convex movement. Locally at least, the relation of convex and linear zones is one of adjacency rather than true interpenetration.

11 Or, as discussed in Chapter 7, will follow long or short models.