Michael Leyton

Group Theory and Architecture

An introduction to a comprehensive theory of design based on group theory in an intuitive form, and building up any needed group theory through tutorial passages. The article will begin by assuming that the reader has no knowledge of group theory, and progressively add more and more group theory in an easy form, until we finally are able to get to quite difficult topics in tensor algebras, and give a group-theoretic analysis of complex buildings such as those of Peter Eisenman, Zaha Hadid, Frank Gehry, Coop Himmelblau, Rem Koolhaas, Daniel Libeskind, Greg Lynn, and Bernard Tschumi. The first part is on a subject of considerable psychological relevance: nested symmetries. In the second Leyton looks at the functional role of symmetry and asymmetry in architecture.

PART I: NESTED SYMMETRIES

Introduction

In my book, *A Generative Theory of Shape* [Leyton 2001], I give a comprehensive theory of design based on group theory. Whereas the book itself requires an advanced knowledge of group theory, the present series of articles, of which this is the first, will give the material in an intuitive form, and build up any needed group theory through tutorial passages. The articles will begin by assuming that the reader has no knowledge of group theory, and we will progressively add more and more group theory in an easy form, until we finally are able to get to quite difficult topics in tensor algebras, and give a group-theoretic analysis of complex buildings such as those of Peter Eisenman, Zaha Hadid, Frank Gehry, Coop Himmelblau, Rem Koolhaas, Daniel Libeskind, Greg Lynn, and Bernard Tschumi. This first article is on a subject of considerable psychological relevance: nested symmetries.

Nested symmetries in architecture

Over the past 20 years, I have been showing in my published research that the human perceptual system is organized as a nested hierarchy of symmetries. This can be demonstrated with many kinds of psychological data on shape perception as well as motion perception. The consequence of the fact that our perceptual systems are organized in this way, is that we structure the environment into nested hierarchies of symmetries. I have also argued that artists and composers exploit this fact in the organization of their works.

The purpose of this article is to introduce the reader to the idea that symmetries in architecture are nested. Let us consider the following example, a colonnade in a cathedral (Figure 1).

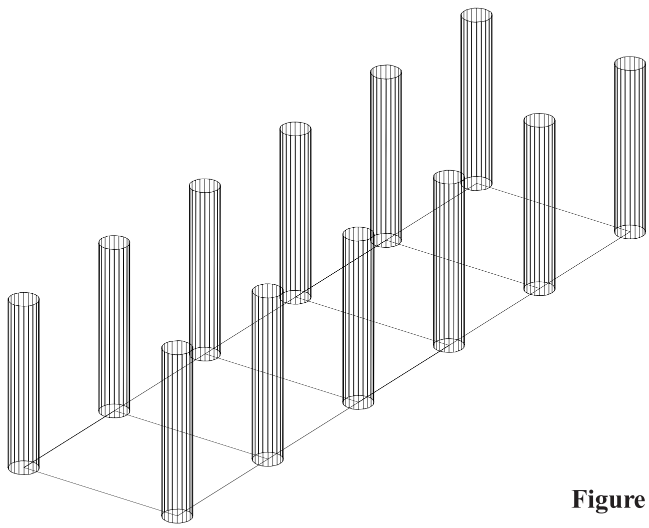


Figure 1

What we will do is build up this structure as a nested hierarchy of symmetries. We will proceed as follows: We start at the lowest level of the organization, which is actually a **Point**. Thus, take a point, and apply to it **Rotations** in the horizontal plane, to generate a circle. This circle is shown on the far left of Figure 2.

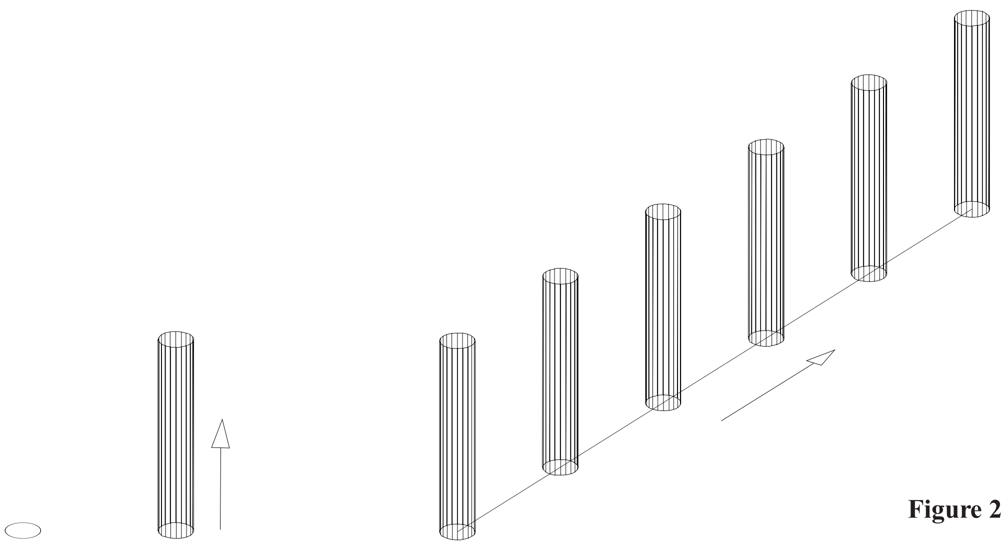


Figure 2

The circle represents the cross-section of the cylinder. Now take the generated circle, and apply to it translations in the vertical direction, which we will denote by $\textbf{Translations}_{\textbf{V}}$. We will get the column, as shown in the second stage in the above figure. The vertical arrow shows the direction of the translations applied to the preceding circle. We next take this column and apply to it $\textbf{Translations}_{\textbf{H}}$ along a line in the horizontal plane as shown on the right in the above figure. This generates a row of columns, as shown in that figure. Finally, we take the row of columns and apply to it **Reflections** about a mirror-plane which is parallel to the column-row, and we obtain the reflectional pair of column-rows shown in figure 1, i.e., the entire colonnade.

The sequence of operations that were used can be represented as follows:

Point. Rotations. Translations$_V$. Translations$_H$. Reflections

where, reading from left to right, we started with a point, then applied rotations to get a column cross-section, then applied vertical translations to get a column, then applied horizontal translations to get a row of columns, and finally applied reflections to get the reflectional pair of column-rows.

The important thing to notice is that these operations were nested. By this I mean the following: Each set of operations generates a level in the architecture. The levels are:
Level 1: A point
Level 2: A circular cross-section
Level 3: A column
Level 4: A row of columns.
Level 5: A pair of column-rows.

Furthermore, each level of transformations acted on the previous level as a whole. This is easy to see as follows: The point was acted on as a whole by **Rotations** to produce a circular cross-section; then the circular cross-section was acted on as a whole by **Translations$_V$** to produce a column; then the column was acted on as a whole by **Translations$_H$** to produce a row of columns; and finally the row of columns was acted on as a whole by **Reflections** to produce the reflectional pair of column-rows.

Each level of transformations defines a symmetry in the architectural structure; i.e., point symmetry, rotational symmetry, translational symmetry, etc. Each level is, in fact, what is called a symmetry group in mathematics.

In my research papers, I call this type of structure 'a hierarchy of nested control'. What I have shown is that the human perceptual system is organized as a hierarchy of nested control. In fact, the first research article I ever published was called "Perceptual organization as nested control." The perceptual system takes its nested structure and imposes it on the environment. What I argue is that architects exploit this psychological fact in the structure of their buildings. But the same is true of painters, and of composers.

Now you might object by saying that the architectural example given above (the colonnade) is a highly regular structure, and therefore amenable to the type of analysis I have given. In contrast, you might ask, how can one describe the new types of architecture that are currently emerging, which involve irregular-shaped blocks (e.g., I.M. Pei's extension to the National Museum in Washington), and also free-form shapes (e.g., Frank Gehry's Guggenheim Museum at Bilbao)? In fact, it was exactly to analyze irregularity and free-form structures that I developed the concept of nested control.

What I have shown is that, given an asymmetric design, the human perceptual system embeds this in a higher dimensional space in which it is described as a nested hierarchy of symmetries. The following is an illustration. In a sequence of psychological experiments

I conducted in the 1980's, I showed that, if people are presented with a rotated parallelogram (far left in Figure 3), they then reference it to a non-rotated one, which they then reference to a rectangle, which they then reference to a square. Thus:

This means that they are actually describing a rotated parallelogram as generated in the following way. One starts with a square (far right). One applies to it a stretch to get a rectangle; then one applies to it a shear to get a parallelogram; and finally one applies to it a rotation to get a rotated parallelogram. This sequence is given thus:

Stretches. Shears. Rotations

Each level is, once again, an example of what mathematicians call a symmetry group. Each is in fact a symmetry of some higher-order space of shapes.

Now we have said that this sequence of operations is applied to a square. However, the square itself is built up as a nested hierarchy of operations. We start with an individual **Point**. We then apply to it **Translations**, to generate a side; and finally we take the side and apply to it the four **90° Rotations** to get a square. That is, the square is described as the following nested hierarchy of symmetries:

Point. Translations. 90° Rotations

Now, we said that the rotated parallelogram is obtained from the square by then applying **Stretches** to get a rectangle; **Shears** to get a parallelogram; and finally **Rotations** to get a rotated parallelogram. So the entire generative sequence, starting with a point, is this:

Point. Translations. 90° Rotations. Stretches. Shears. Rotations

The first three operations produce the square successively from a point, and then the next three operations produces the rotated parallelogram successively from a square.

It turns out that this 6-level structure is a hierarchy of nested symmetries. Each level is a symmetry of some space, and the spaces are nested in each other. This 6-level nested hierarchy of control is a very powerful structure in the human perceptual system. I have shown, for example, that it structures not only geometrical figures, such as those given above, but also motion phenomena.

Now, we can go on adding higher levels of operations which will make the shape more and more asymmetric. As an example, I invented and published a free-form grammar which alters the curvature of the shape, so that it takes on more and more of an organic growth appearance. The grammar exactly analyzes, for example, Gehry's Guggenheim museum.

Extensive research on this topic has been published in *Symmetry, Causality, Mind* [Leyton 1999], which contains many examples of how the human mind structures the environment in terms of nested hierarchies of symmetries. The first two chapters of the book give the free-form grammar that I just mentioned. Also included is a description of the application of the free-form grammar to analyzing actual artworks in detail. The book takes the same intuitive approach as the present paper.

In contrast, all the technical mathematics will appear in *A Generative Theory of Shape*, to be published by Springer-Verlag in September 2001.

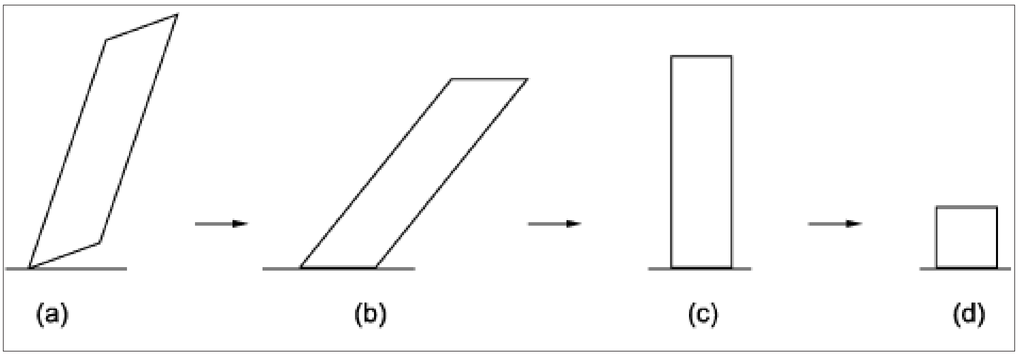


Figure 3

PART II: WHY SYMMETRY / ASYMMETRY?

Introduction

In this second tutorial we are going to look at the *functional role of symmetry and asymmetry in architecture.* We are all aware that classical architecture was dominated by symmetry. In contrast, we have seen, in the twentieth century, a shift from the dominating role of symmetry to the gradual raising of asymmetry as the major principle. Famous examples of the latter include Frank Lloyd Wright's Falling Water, with its asymmetrically arranged blocks, or Eero Saarinen's TWA Building with its free form structure, or in the contemporary world, the Deconstructivist Architects are now the dominant force. The latter movement came into significant public recognition with the exhibition of their work in the Museum of Modern Art, New York, in1988, and these architects are now the most famous archictects in the world— usually winning the major architecture competitions. They include Peter Eisenman, Zaha Hadid, Frank Gehry, Coop Himmelblau, Rem Koolhaas, Daniel Libeskind, and Bernard Tschumi. In all their buildings, *asymmetry* is the major organizing factor.

What we wish to consider, in this paper, is the following issue: Why was classical architecture dominated by symmetry; i.e., what purpose did symmetry serve in classical architecture? Correspondingly, why is modern architecture dominated by asymmetry; i.e., what purpose does asymmetry serve in modern architecture?

The answer to this question comes from my previous book *Symmetry, Causality, Mind* [1999], in which I argue that symmetry is always used to erase memory from an organization, and asymmetry is always used to introduce memory into an organization. I show that these memory principles are deeply embedded in the human mind: indeed they are what allows the mind to work. It is these memory principles, I argue, that are at the basis of classical architecture's use of symmetry and the modern architecture's use of asymmetry. That is, classical architecture is aimed at removing memory, and contemporary architecture aims at creating memory.

Inferring history from shape

The book *Symmetry, Causality, Mind* [Leyton 1998] presents a 630-page rule-system by which the mind extracts the past history that produced a shape, i.e., the sequence of causal forces that produced the shape. Despite the enormous number of rules they all are different forms of only two basic rules: one that exploits the asymmetries in a shape, and one that exploits the symmetries in the shape. The theory ultimately explains how any organization can hold "memory" of past actions.

If we define "memory" to be information about the past, we observe that there are many forms that memory can take. For example, a *scar* is memory of past events because, when we look at it, we are able to extract *information about past actions*, i.e., the fact that there had previously been a past cutting action across the skin. Again, a *crack* in a vase is memory of past events because, when we look at it, we are able to extract *information about past actions*, i.e., the fact that there had previously been a blow applied to the vase. There are in fact an almost infinite number of forms that memory can take: scars, cracks, dents, twists, growths, and so on. However arguments presented in my book [Leyton 1992], lead to the conclusion that, on an abstract level, there is only one form that memory takes:

Memory is always in the form of asymmetry.

Symmetry is always the absence of memory.

I can give you a simple illustration of this as follows: Imagine a tank of gas on the table. Imagine that the gas is at equilibrium, at TIME 1. The gas is therefore uniform throughout the tank, in particular, symmetric—left to right in the tank. Now use some means to attract the gas into the left half of the tank at TIME 2. The gas is now asymmetric.

Someone who has not previously been in the room now enters and sees the gas. The person will immediately conclude that the gas underwent a movement to the left. This means that the asymmetric state is memory of the movement. Now let the gas settle back to equilibrium, that is symmetry at TIME 3, that is, uniformity throughout the tank.

Suppose another person enters now, someone who has not been in the room before. This new person would not be able to deduce that the gas had previously moved to the left and returned. The reason is that the symmetry has wiped out the memory of the previous events. The conclusion is that from symmetry, you can conclude only that the past was the same. We can summarize the rules used here, in two principles:

ASYMMETRY PRINCIPLE:

An asymmetry in the present is assumed to have been a symmetry in the past.

SYMMETRY PRINCIPLE:

A symmetry in the present is assumed to have always existed.

In mathematics, symmetry means indistinguishability under transformations. Thus, for example, a face is reflectionally symmetric because it is indistinguishable from its reflected version, and a circle is rotationally symmetric because it is indistinguishable from any of its rotated versions.

Now, what we will see over and over again in this paper is that the way to use the above two rules is as follows: You first partition the present situation into its asymmetries and symmetries. You then use the first rule on the asymmetries and the second rule on the symmetries. That is, the first rule says that the asymmetries go to symmetries, backward in time; and the second rule says that the symmetries are preserved, backward in time.

Let us now illustrate this: In a converging series of psychological experiments, I showed that if subjects are presented with the first stimulus shown in Figure 3 above, a rotated parallelogram, they reference it in their minds to a non-rotated parallelogram, which they then reference to a rectangle, which they then reference to a square. The important thing to understand is that they are presented with only the first figure; and, from this, their minds generate the sequence shown.

One can interpret this data by saying that, given the initial object, subjects are inferring the process-history that produced it. That is, the presented object was produced by starting with a square, stretching it, then shearing it, and then rotating it.

We shall now see that what the subjects are doing is using the Asymmetry Principle and Symmetry Principle. To see this, we must, as I said, first partition the presented shape—the rotated parallelogram—into its asymmetries and its symmetries. Consider first the asymmetries. There are in fact three of them: (1) the distinguishability between the orientation of the shape and the orientation of the environment; (2) the distinguishability between adjacent angles; (3) the distinguishability between adjacent sides.

As we can see from Figure 3, what subjects are doing is removing these three distinguishabilities, backwards in time as prescribed by the Asymmetry Principle. That is, successively, the orientation of the shape becomes the same as that of the environment, the sizes of the adjacent angles becomes the same, and the sizes of the adjacent sides become the same. To repeat: Asymmetries become symmetries backward in time—as predicted by the Asymmetry Principle.

Now let us use the Symmetry Principle. It says that the symmetries must be preserved, backward in time. Well, the rotated parallelogram has two symmetries: (1) opposite angles are indistinguishable in size; and (2) opposite sides are indistinguishable in length. Observe that both of these symmetries are preserved backward in time—thus corroborating the Symmetry Principle.

Now, those of you who have seen my book might say to me, "There seem to be hundreds of rules in your book. How can you say that there are actually only two rules?" Well, the reason is that, as I said earlier, the term symmetry means indistinguishability under transformations: reflectional symmetry is indistinguishability under reflectional transformations; rotational symmetry is indistinguishability under rotational transformations, and so on. Thus you obtain the different kinds of symmetry by instantiating the different kinds of transformations in the definition of symmetry. The different rules of the book are obtained by instantiating different transformations within the Asymmetry Principle and Symmetry Principle. Notice that it is by doing this instantiation process that you obtain the different sources of memory that can exist in an organization.

In the paper so far, I have given you only an intuitive sense of the instantiation process. What I want to do now is show you how it works, in depth. We are going to examine the extraction of memory from a particular asymmetry called *curvature extrema*. We will see later that curvature extrema are violations of rotational symmetry in the outline of a shape.

So lets look at curvature extrema. What is a curvature extremum? Well, first we note that curvature, for curves in the 2D plane, is simply the amount of bend. The straight line has no bend, and therefore has no curvature. As you successively increase bend, you are increasing curvature. Finally, observe that on a shape such as a finger, there is a point that has more bend than the other points on the line (the finger tip). This is a curvature extremum.

We will start be elaborating two successive rules by which the curvature extrema can be used to infer processes that have acted upon a shape. The input to the rules will be smooth outlines of shapes such as embryos, tumors, clouds, etc. So the rules will infer the history of such objects — that is, convert them into memory.

The inference, from curvature extrema to historical processes will be seen as requiring two stages: (1) Curvature extrema Symmetry axes, and (2) Symmetry axes Processes. We first consider stage 1.

The PISA symmetry analysis

As we have seen, an essential aspect of the inference of history is symmetry. For a simple shape, a symmetry axis is usually defined to be a straight line along which a mirror will reflect one half of the figure onto the other. However, observe that, in complex natural objects, such as the branch of a tree, a straight axis might not exist. Nevertheless, one might still wish to regard the object, or part of it, as symmetrical about some *curved axis*. For example, a branch of a tree tends not to have a straight reflectional axis. Nevertheless, one understands the branch to have an axial core that runs along its center.

How can such a *generalized* axis be constructed? There are good mathematical reasons to proceed in the following way. Consider the next figure, Figure 4. It shows two curves c_1 and c_2 (the bold curves), which can be understood as two sides of an object. We will construct a symmetry axis between the two curves. In fact, the axis we will produce will be indicated by the sequence of dots shown in Figure 4. It is produced in the following way: introduce a circle that is tangential simultaneously to the two curves, as shown; the two tangent points A and B are then defined as symmetrical to each other.

Then move the circle continuously along the two curves, c_1 and c_2, while always maintaining the property that it is simultaneously tangential to the two curves. One might need to expand or contract the circle to maintain the double-touching property. Now define the symmetry axis to be the trajectory of the midpoint Q of the arc AB, as the circle moves. For example, in the above figure, the trajectory of Q is represented by the locus of dots shown. This definition of symmetry was proposed by me in Leyton [1988], where I argued that it is more suited for the particular task of process-inference than are other symmetry analyses, such as those proposed by Blum and Brady. The

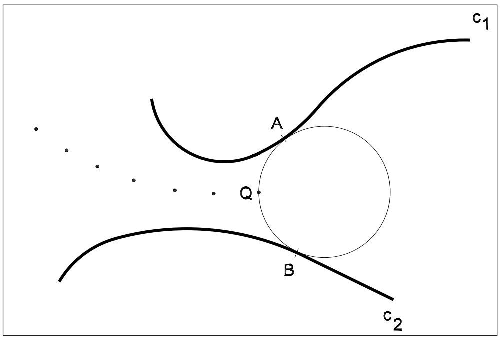


Figure 4

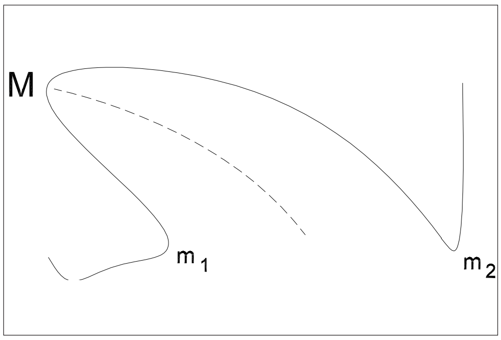


Figure 5

intuitively satisfying results of our alternative analysis will be seen later. I called the above definition of symmetry axis, *Process-Inferring Symmetry Axis*, or simply *PISA*.

Symmetry-curvature duality

The Free-Form Grammar relies on two structural factors in a shape: symmetry and curvature. Mathematically, symmetry and curvature are two very different descriptors of shape. However, a theorem that I proposed and proved in Leyton [1987b] shows that there is an intimate relationship between these two descriptors. This relationship will be the basis of the entire paper:

SYMMETRY-CURVATURE DUALITY THEOREM

Any section of curve that has one and only one curvature extremum has one and only one symmetry axis. This axis is forced to terminate at the extremum itself.

To illustrate: Consider the shape shown in Figure 5. The section of curve between the two letters *m*, has only one curvature extremum —that indicated by the letter *M*. The theorem says that this section of curve can have only one symmetry axis, and that the axis is forced to terminate at the extremum.

Symmetry axes and processes

The reason for involving symmetry axes is that it will be argued that they are closely related to *process-histories*. This proposed relationship is given by the following principle [Leyton 1984]:

INTERACTION PRINCIPLE

The symmetry axes of a perceptual organization are interpreted as the directions along which processes are most likely to act or have acted.

The principle was advanced and extensively corroborated in Leyton [1984, 1985, 1986a, 1986b, 1986c, 1987a, 1987b, 1987c], in several areas of perception including motion perception as well as shape perception. The argument used in Leyton [1984, 1986b] to justify the principle involves the following two steps:
(1) A process that acts along a symmetry axis tends to preserve the symmetry; i.e. to be

structure-preserving;
(2) Structure-preserving processes are perceived as the most likely processes to occur or to have occurred.

The inference of processes

We now have the tools required to understand how processes are recovered from shape. In fact, the system to be proposed consists of two inference rules that are applied successively to a shape. The rules can be illustrated considering Figure 6.

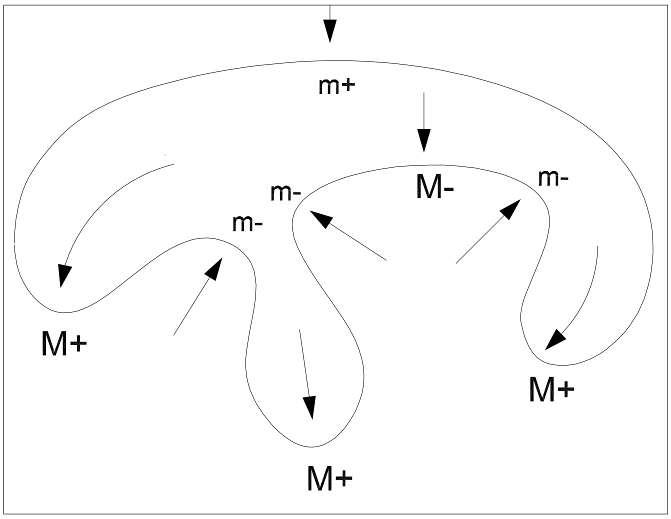


Figure 6

The first rule is the Symmetry-Curvature Duality Theorem, which states that to each curvature extremum, there corresponds a unique symmetry axis terminating at that extremum. The second rule is the Interaction Principle, which states that each of the axes is a direction along which a process has acted. The implication is that the boundary was deformed along the axes; i.e., each protrusion was the result of *pushing out* along its axis, and each indentation was the result of *pushing in* along its axis. In fact, each axis is the *trace* or *record* of boundary-movement! (The vertical arrow at the top is a *squashing* and the rules for inferring this are the same, as we shall see later.)

Under this analysis, processes are understood as creating the curvature extrema; i.e., the processes introduce protrusions, indentations, etc. into the shape boundary. This means that, if one were to go backwards in time, undoing all the inferred processes, one would eventually remove all the extrema. Observe that there is only one closed curve without extrema: the circle. Thus the implication is that the ultimate starting shape must have been a circle, and this was deformed under various processes each of which produced an extremum.

Corroborating examples

To see that our two rules consistently yield satisfying process-explanations, let us obtain the processes that these rules give for a large set of shapes. We shall take the set of all possible shapes that have 8 extrema or less. Such shapes fall into three levels: shapes with 4 extrema, shapes with 6 extrema, and shapes with 8 extrema. The reason is that there cannot be shapes with an odd number of extrema — because maxima have to alternate with minima of curvature — and there cannot be shapes with less than 4 extrema by a theorem in differential geometry. We will refer to the three successive levels of shapes as Levels 1, 2, and 3 respectively. There are a total of 21 shapes, of succesively increasing complexity.

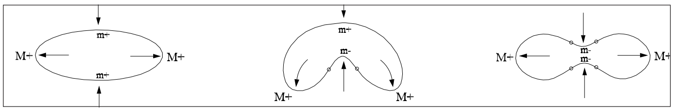


Figure 7

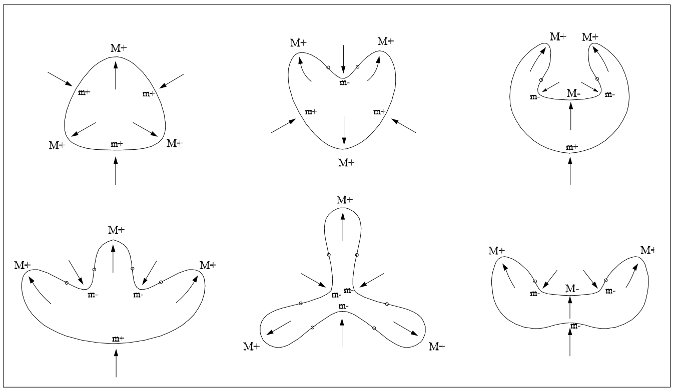


Figure 8

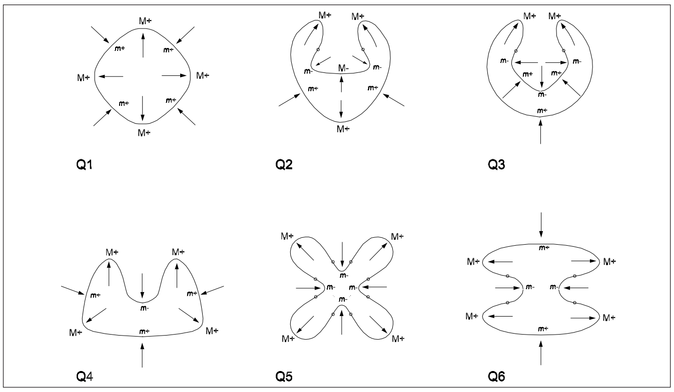


Figure 9

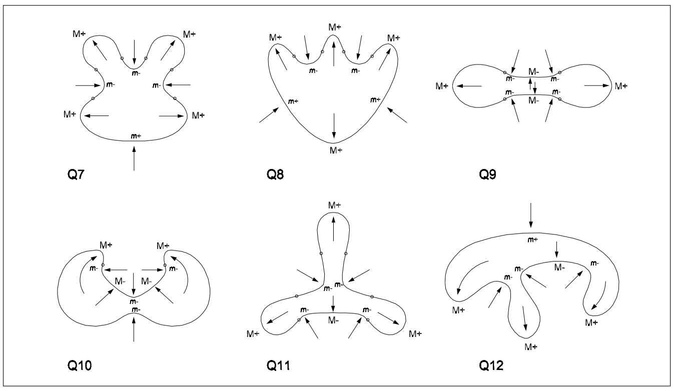


Figure 10

Let me make the following comments:

(1) The reader will notice that, on each shape, each extremum is marked by one of four symbols: M+, M-, m+, m-. This is because there are mathematically four kinds of curvature extrema: Positive Maxima (M+); Negative Maxima (M-), Positive minima (m+); and Negative minima (m-).

(2) When one surveys the shapes, one finds that there is the following simple rule that relates the type of extremum to an English word for a process:

SEMANTIC INTERPRETATION RULE

M+ is always a *protrusion*
m- is always an *indentation*
m+ is always a *squashing*
M- is always an *internal resistance*

Remember as you look at the figures that the process arrows are inferred by the two simple rules we gave above:
The Symmetry-Curvature Duality Theorem,
and the Interaction Principle.

The 21 shapes are shown in the following figures: Shapes with 4 extrema: (Figure 7); Shapes with 6 extrema: (Figure 8); Shapes with 8 extrema: Set 1 (Figure 9); Set 2 (Figure 10).

Let us now go more deeply into the structure of these histories.

The free-form grammar

Up to now we have considered the issue of how one infers processes from a *single* shape. We will now examine a different problem: Supposing one is given *two* views of an object (e.g., an embryo), at two different stages of development. How can one infer the *intervening* process-history? Our method of solving this problem will be to develop a *grammar* that generates the second shape from the first via a sequence of plausible developmental stages.

Observe now that, since the later shape is assumed to emerge from the earlier shape, one will wish to explain it, as much as possible, as the outcome of what can be seen in the earlier shape. In other words, one will wish to explain the later shape, as much as possible, as the *extrapolation* of what can be seen in the earlier one.

As a simple first cut, let us divide all extrapolations of processes into two types:

(1) Continuations
(2) Bifurcations (i.e., branchings).

What we will do now is elaborate the only forms that these two alternatives can take. We first look at continuations and then at bifurcations.

Continuations

Consider any one of the *M*+ extrema in Figure 11. It is the tip of a protrusion, as predicted by our Semantic Interpretation Rule. What is important to observe is that, if one continued the process creating that protrusion, i.e. continued pushing out the boundary in the direction shown, the protrusion would remain a protrusion. That is, the *M*+ extremum would remain a *M*+ extremum. This means that continuation at a *M*+ extremum does not structurally alter the boundary.

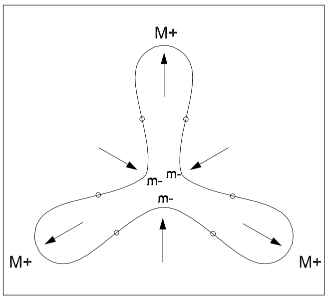
Exactly the same argument applies to any of the *m*- extrema in the above figure. That is, continuation of an indentation remains an indentation. That is, a *m*- extremum will remain *m*-

Now recall that there are four types of extrema, *M*+, *m*-, *m*+, *M*-. We have seen that continuations at the first two do not structurally alter the shape. However, we shall now see that continuations at the second two do cause structural alteration. Let us consider these two cases in turn.

CONTINUATION AT *m*+

A *m*+ extremum occurs at the top of the left-hand shape in Figure 12. In accordance with our Semantic Interpretation Rule, the process terminating at this extremum is a *squashing* process; i.e., it explains the flattening at the top of the shape. Now let us continue this process; i.e., continue pushing the boundary in the direction shown. At some point, an indentation will be created, as shown in the top of the right-hand shape in Figure 13.

Observe what happens to the extrema involved. Before continuation, in the left-hand shape, the relevant extremum is *m*+ (at the top). After continuation, in the right-hand



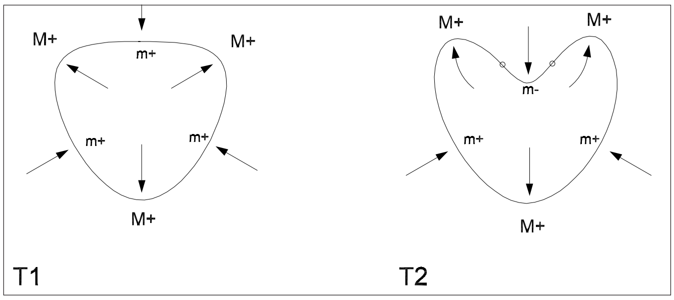


Figure 11 **Figure 12**

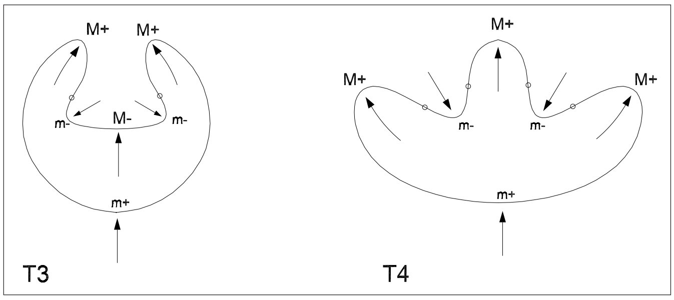


Figure 13

shape, this extremum has changed to *m-*. In fact, observe that a dot has been placed on either side of the *m-*, on the curve itself. These two dots are points where the curve, locally, is completely flat; i.e., where the curve has 0 curvature. Therefore the top of the right-hand shape is given by the sequence *0 m- 0*. The transition between the left-hand shape and the right-hand shape can be structurally specified by simply saying that the *m+* extremum, at the top of the first shape, is replaced by the sequence *0 m- 0*, at the top of the second shape. This transition will be labeled *Cm+*, meaning **Continuation at *m+***. That is, we have:

$$Cm+ : m+ \rightarrow 0\ m\text{-}\ 0$$

The above string of symbols says "Continuation at *m+* takes the *m+* and changes it into the triple *0 m- 0*".

Observe that, although this operation is, formally, a rewrite rule on discrete strings of extrema, the rule actually has a highly intuitive meaning. Using our Semantic Interpretation Rule, we see that it means: *squashing continues till it indents*.

CONTINUATION AT *M-*

As noted earlier, we need to consider only one other type of continuation, that at a *M-* extremum. In order to understand what happens here, consider the left-hand shape in Figure 13. The symmetry-analysis given at the beginning of this paper describes a process-structure for the indentation that is very subtle, as shown: there is a flattening of the lowest region of the indentation due to the fact that the downward arrows, within the indentation, are countered by an upward arrow, which is within the body of the shape and terminates at the *M-* extremum. This latter process is an example of what our Semantic Interpretation Rule calls *internal resistance*. The overall shape could be that of an island where an inflow of water (into the indentation) has been resisted by a ridge of mountains (in the body of the shape). The consequence is the formation of a *bay*.

Now, recall that our interest is to see what happens when one continues a process at a *M-* extremum. Thus let us continue the *M-* process upward, in the left-hand shape in Figure 11. At some point, the process will burst out and create the protrusion shown at the top of the right-hand shape. In terms of our island example, there might have been a volcano, in the mountains, that erupted and sent lava down into the sea.

Now let us observe what happens to the extrema involved. Before continuation, in the left-hand shape, the relevant extremum is *M-* (in the center of the bay). After continuation, in the right-hand shape, this extremum has changed to *M+* (at the top of the protrusion). In fact, observe that, once again, a dot has been placed on either side of the *M+*, on the curve itself. These two dots represent, as before, points where the curve is, locally, completely flat; i.e., where the curve has 0 curvature. Therefore the top of the right-hand shape is given by the sequence *0 M+ 0*.

Therefore the transition between the left-hand shape and the right-hand shape can be structurally specified by simply saying that the *M-* extremum, in the left-hand shape, is replaced by the sequence *0 M- 0*, at the top of the right-hand shape. This transition will be labeled *CM-*, meaning **Continuation at *M-***. Thus we have:

$$CM\text{-} : M\text{-} \rightarrow 0\ M\text{+}\ 0$$

The above string of symbols says "Continuation at *M-* takes the *M-* and changes it into the triple *0 M+ 0.*"

Observe once again, that, although this operation is a formal rewrite rule on discrete strings of extrema, the rule actually has a highly intuitive meaning. Using our Semantic Interpretation Rule, we see that it means: *internal resistance continues till it protrudes.*

Bifurcation

We saw above that process-continuation can take only two forms. Our purpose now is to elaborate the only forms which the *bifurcation* (branching) of a process can take. Note that, because there are four extrema, we have to examine bifurcation at each of these four.

BIFURCATION AT *M+*

Consider the *M+* extremum at the top of the left-hand shape in Figure 14, and consider the upward protruding process terminating at this extremum. We wish to examine what would result if this process bifurcated. Under bifurcation, one branch would go to the left and the other to the right. That is, the branching would create the upper lobe in the right-hand shape in Figure 14.

Observe what happens to the extrema involved. Before splitting, one has the *M+* extremum at the top of the first shape. In the situation after splitting (i.e. in the second shape), the left-hand branch terminates at a $M+$ extremum, and so does the right-hand branch. That is, the $M+$ extremum, in the first shape, has split into two copies of itself in the second shape. In fact, for mathematical reasons, a new extremum has to be introduced in between these two $M+$ copies. It is the $m+$ shown at the top of the lobe of the second shape. Therefore, in terms of extrema, the transition between the first and second shapes can be expressed thus: The $M+$ extremum at the top of the first shape is replaced by the sequence $M+\ m+\ M+$ along the top of the second shape. This transition will be labeled $BM+$, meaning **Bifurcation at *M+***. That is, we have:

$$BM+ : M+ \rightarrow M+\ m+\ M+$$

The above string of symbols says "Bifurcation at $M+$ takes the $M+$ and changes it into the triple $M+\ m+\ M+$."

Observe that, although this transition has just been expressed as a formal re-write rule on discrete strings of extrema, the transition has, in fact, the following highly intuitive meaning: *a nodule develops into a lobe*.

BIFURCATION AT *m-*

Consider now the $m-$ extremum at the top of the left-hand shape in Figure 15, and consider the downward indenting process terminating at this extremum. We will examine what results if this process bifurcates. Under bifurcation, one branch would go to the left and the other to the right. That is, the branching would create the *bay* in the right-hand shape.

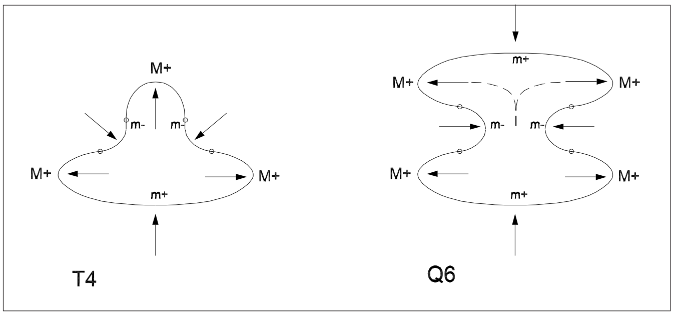


Figure 14

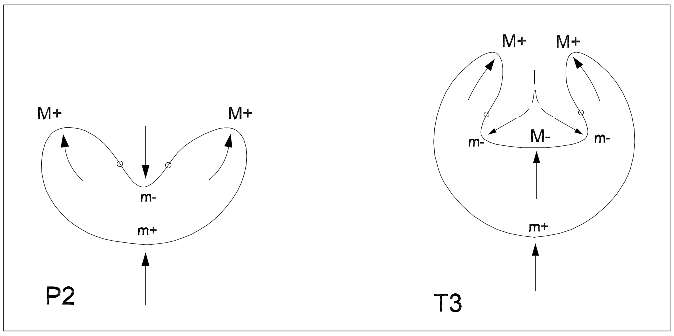


Figure 15

Observe, again, what happens to the extrema involved. Before splitting, one has the single extremum $m-$ in the indentation of the first shape. In the situation after splitting (i.e., the second shape), the left-hand branch terminates at a $m-$ extremum, and so does the right-hand branch. That is, the top $m-$ in the first shape has been split into two copies of itself, in the second shape. Again, for mathematical reasons, a new extremum has to be introduced in between these two *m-* copies. It is the *M-* shown at the center of the bay in the second shape. Therefore, in terms of extrema, the transition between the first and second shapes can be expressed thus: The *m-* extremum in the first shape is replaced by the sequence *m- M- m-* in the second shape. This transition will be labeled *Bm-*, meaning **Bifurcation at *m-***. That is, we have:

$$Bm\text{-} : m\text{-} \rightarrow m\text{-}\ M\text{-}\ m\text{-}$$

The above string of symbols says "Bifurcation at *m-* takes the *m-* and changes it into the triple *m- M- m-*."

Observe that, although the transition has just been expressed as a formal re-write rule using discrete strings of extrema, the transition has, in fact, a highly intuitive meaning: *an inlet develops into a bay*.

BIFURCATION AT *m+* and *M-*

Bifurcations at these two extrema turn out to be very easy to understand: They are simply the introduction of a protrusion and the introduction of an indentation, respectively. It will therefore not be necessary to diagram them.

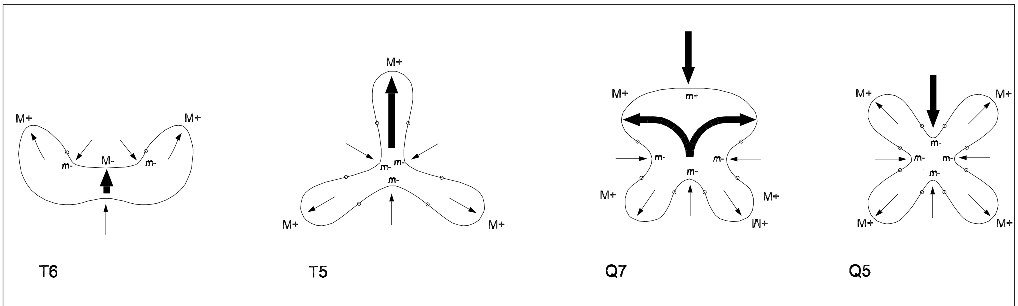


Figure 16

The complete grammar

Recall that the problem we are examining is this: Given two views of an object (e.g., a tumor), at two different stages of development, how is it possible to infer the *intervening shape-evolution*? Observe that, since one wishes to explain the later shape as an outcome of the earlier shape, one will try to explain the later shape, as much as possible, as the *extrapolation* of the process-structure inferrable from the earlier shape. We have now shown that all possible process-extrapolations are generated by only six operations: two continuations and four bifurcations. Therefore, these six operations form a grammar that will generate a later shape from an earlier one via process-extrapolation. The grammar is as follows:

FREE-FORM GRAMMAR

Cm+ : *m+* → 0 m- 0

CM- : *M-* → *0 M+ 0*

BM+ : *M+* → *M+ m+ M+*

Bm- : *m-* → *m- M- m-*

Bm+ : *m+* → *m+ M+ m+*

BM- : *M-* → *M- m- M-*

Recall however that, although these operations are expressed as formal re-write rules on discrete strings of extrema, they describe six intuitively compelling situations, as follows:

SEMANTIC INTERPRETATION OF THE GRAMMAR

Cm+ : squashing continues till it indents.
CM- : internal resistance continues till it protrudes.
BM+ : a protrusion bifurcates; e.g. a nodule becomes a lobe.
Bm- : an indentation bifurcates; e.g. an inlet becomes a bay.
Bm+ : a protrusion is introduced.
BM- : an indentation is introduced.

These situations are illustrated in the previous figures.

An illustration of shape evolution

Figure 16 shows an illustration of shape evolution. Each successive stage is given by an operation of the grammar. The sequence of operations is *CM-* followed by *BM+* followed by *Cm+*.

Let us describe what his happening in Figure 16:

(1) The entire history is dependent on a single crucial process. It is the *internal resistance* represented by the bold upward arrow in the first shape.

(2) This process continues upward until it bursts out and creates the protrusion shown in the second shape.

(3) This process then bifurcates, creating the upper lobe shown in the third shape above. Involved in this stage is the introduction of a squashing process given by the bold downward arrow at the top of the third shape.

(4) Finally, the downward squashing process continues till it causes the new indentation in the top of the fourth shape above.

Back to symmetry and asymmetry

This concludes my exposition of the rules for obtaining history from curvature extrema. There are a total of eight rules:

EXTREMA-BASED RULES
1. Symmetry-Curvature Duality Theorem
2. Interaction Principle
3-8. The six rules of the Free-Form Grammar.

I will refer to these rules as the extrema-based rules. They pick certain features—the curvature extrema—and extract causal history from those features; that is, they construct memory from those features.

What I want to do now is show you that these eight rules are an instantiation of the theory of process-inference, or memory construction, that was given in the first part of the paper. Before I do this, it is necessary to understand first that curvature variation is a form of rotational asymmetry. To understand this, imagine that you are driving a car on a racing track which is in the shape of one of the curvilinear shapes I showed you. On such a track, there is no alternative but to keep on adjusting the steering wheel as you are driving. This is because curvature is changing at all points. In contrast, if you are driving on a track that is perfectly circular, you would have to set the wheel only once, at the beginning, and never have to adjust it again. This is because the curvature is the same at all points on a circle. Another way of saying this is that a circle is rotationally symmetric—that is, in going around the circle, each section is indistinguishable from any other. Therefore, we can now see that what curvature extrema do is introduce rotational asymmetry in the shape.

So much for the asymmetry in a curvilinear shape. The Asymmetry Principle will be applied to this asymmetry and remove it backward in time. What about the symmetry in such a shape. The Symmetry Principle will be applied to that and preserve it backward in time. Observe that despite all the asymmetry in such a shape, there is a form of symmetry. It is reflectional symmetry. It is exactly captured by the symmetry axes.

Now, having understood this, let us now go through our eight rules, and see how they are an instantiation of our scheme for process-inference, or memory-construction. What was our scheme? It was in fact illustrated several times in the paper already: You first partition the situation into its asymmetry and symmetry components, and then you apply the Asymmetry Principle to the asymmetry component, and the Symmetry Principle to the symmetry component.

Let us now do through the eight rules and show how each is designed to carry out a role in this scheme: First we have the Symmetry-Curvature Duality Theorem. This theorem corresponds each curvature extremum to a symmetry axis. We can now understand that what the theorem is now doing is, in fact, describing the exact relationship between the asymmetry component and the symmetry component. It says that, for each unit of asymmetry, that is, for each curvature extremum, you will find a unit of symmetry, a symmetry axis. In other words, the role of the theorem is to carry out the initial partitioning stage in the inference process.

Now let us look at the next rule. It is the Interaction Principle, which says that processes have to have gone along the symmetry axes. This has the effect of preserving the symmetry axes over time. In fact, brief consideration reveals that this principle is merely an example of the Symmetry Principle — the injunction to preserve symmetries backwards in time.

Now lets move onto the six rules of the Free-Form Grammar. What do these six rules do? The answer is that they describe the six only possible ways in which curvature variation can increase in a shape. In other words, they are the six only possible instantiations of the Asymmetry Principle when the asymmetry is curvature variation.

So now we can understand exactly how this entire system of eight rules instantiates our scheme for the extraction of memory: The Symmetry-Curvature Duality Theorem specifies the partitioning of the shape into its asymmetry and symmetry components; the Interaction Principle is an instantiation of the Symmetry Principle; and the Free-Form Grammar is an instantiation of the Asymmetry Principle.

What I have done in this part of the paper is shown just one of the rule-systems that I developed in my book *Symmetry, Causality, Mind.* In fact, this system is given right at the beginning of the book, and after that there are another 600 pages of rule systems. Each of these systems is applied to a different type of structural feature in an organization to extract causal history from that feature. Since I published them, my rule-systems have been applied by scientists in many disciplines to extract causal history, e.g., meteorology, radiology, chemical engineering, linguistics, etc.

Back to architecture

Let us now return to the topic of symmetry and asymmetry in architecture. We can see, from the above discussion, that the Asymmetry Principle and Symmetry Principle lead to the following conclusions:

Classical architecture aimed at removing memory.

Contemporary architecture aims at creating memory.

In addition to these general principles, we have the particular rules of the Free-Form Grammar. This will allow us, in future articles, to do careful analyses of Frank Gehry's Guggenheim Museum at Bilbao as well as the free-form buildings of Greg Lynn. Furthemore, our other memory rule-systems, derived from the Asymmetry Principle and Symmetry Principle, will enable us to analyze the non-free-form buildings of the Deconstructivist Architects, and Lebbeus Woods.

Acknowledgment

These papers were previously published in *VisMath*, vol. 1 no. 3 and no. 4 (1999) and are reproduced by permission, courtesy of Slavik Jablan.

References

Leyton, M. 1984. Perceptual organization as nested control. *Biological Cybernetics* **51**: 141-153.

———. 1986a. Principles of information structure common to six levels of the human cognitive system. *Information Sciences* **38**: 1-120. (Entire journal issue.)

——— . 1986b. A theory of information structure I: General principles. *Journal of Mathematical Psychology* **30**: 103-160.

——— . 1986c. A theory of information structure II: A theory of perceptual organization. *Journal of Mathematical Psychology* **30**: 257-305.

——— . 1987a. *Nested structures of control: An intuitive view. Computer Vision, Graphics, and Image Processing* **37**: 20-53.

——— . 1999. *Symmetry, Causality, Mind*. Cambridge MA: MIT Press.

——— . 2001. *A Generative Theory of Shape.* Berlin: Springer-Verlag.

About the author

Michael Leyton obtained his BSc with Honours in Mathematics from Warwick University, England in 1977, and his Ph.D from the Psychology Department, Berkeley, CA, in 1984. After a post-doctoral research fellow at Harvard University, in 1986 he became Assistant Professor of Computer Science and Psychology at SUNY. In 1987 he joined Rutgers University where in 1993 he became Associate Professor of Psychology. He has been awarded of some important prizes (1987-1992, Presidential Young Investigator Award in Cognitive Science; 1987-1989, HRRF Award; 1988, Medal of Achievement of the Society for Applied Systems Research). He is President of The International Society for Group Theory in Cognitive Science, and of The International Society for Mathematical and Computational Aesthetics. Dr. Leyton is author of *Symmetry, Causality, Mind* (MIT Press, 1992). His new book is *A Generative Theory of Shape* (Springer-Verlag, 2001). Dr. Leyton has written important papers on shape perception and has been invited speaker in many international conferences.