Environment and Planning B: Planning and Design 1997, volume 24, pages 359–384

Spatial change: continuity, reversibility, and emergent shapes

R Krishnamurti
Department of Architecture, Carnegie Mellon University, Pittsburgh, PA 15213,USA;
e-mail: ramesh@arc.cmu.edu
R Stouffs
Architecture and CAAD, Swiss Federal Institute of Technology, Rämistr. 101, ETH-Zentrum, 8092 Zurich, Switzerland
Received 28 November 1995; in revised form 9 April 1996

Abstract. Spatial composition can be viewed as computations involving spatial changes each expressed as $s - \mathrm{f}(a) + \mathrm{f}(b)$, where s is a shape, and $\mathrm{f}(a)$ is a representation of the emergent part (shape) that is altered by replacing it with the shape $\mathrm{f}(b)$. We examine this formula in three distinct but related ways. We begin by exploring the conditions under which a sequence of spatial changes is continuous. We next consider the conditions under which such changes are reversible. We conclude with the recognition of emergent shapes, that is, the determination of transformations f that make $\mathrm{f}(a)$ a part of s. We enumerate the cases for shape recognition within algebras U_{ij}, $0 \leqslant i \leqslant j \leqslant 3$, and within Cartesian products of these algebras.

Spatial change

Computer-aided design (CAD) is often nothing more than an euphemism for computer-aided drafting, generally referring to systems that serve as repositories for *designed* information. All too often newer approaches disguise old technology in new garments.[1] Design is about change; design systems are formalisms which accommodate a notion of change. CAD is a process that employs computational mechanisms to effect change. The manner of change affects the way the worlds of possible designs can be explored. A particularly enticing and related concept is that of emergence (Mitchell, 1993; Stiny, 1993a; 1994).

New spatial objects and changes to spatial objects are produced by manipulative operations on spatial objects. We consider a system for manipulating such objects. Such a system (at least, implicitly) includes all objects that can be produced by the manipulative operations, which at a minimum, consist of the spatial arithmetic operations and geometrical transformations. Let s be an object (of interest) within this system. The basic operation of producing a new shape (to the system) is by 'adding' to s in a specified manner, which we can describe as $s \rightarrow s + b$. The symbol $\rightarrow$ denotes a derivation of a new spatial object from a given object. In this case, the new object is derived from the given object, s, by adding the object b. We may of course add the spatial object b, not as simply but through some functional or transformational form of it. Thus, $s \rightarrow s + \mathrm{f}(b)$, where $\mathrm{f}(b)$ is some transformation of the object b.

[1] To some, taken literally, this statement is patently false when applied to disciplines where the foci of interest centre on the functional and behavioural descriptions associated with designed objects. These descriptions are (typically, hierarchically) structured relationships between items of known semantic types. In disciplines where design is through assembly, object-oriented approaches or case-based design can be useful and effective. Where design proceeds by composition, where the objective is not solely the designed object but also the discovery or manipulation of new 'components' along the way, these methods, by their very nature, are ineffectual. It is this aspect of novelty which we believe to be fundamental to design.

Equally, a new object can be produced by 'removing' from s in some specific manner, which we can describe as $s \rightarrow s - \mathrm{f}(a)$. We note here that the production of new spatial objects by updating some aspects or properties of an existing object can be regarded as 'addition' or 'subtraction' depending on how the function f and the operators '+' and '−' are defined.

We can of course effect change to s by the removal and subsequent addition of aspects or properties of s by the formula $s - \mathrm{f}(a) + \mathrm{g}(b)$, where, in effect, $\mathrm{f}(a)$ is the altered part and $\mathrm{g}(b)$ is its replacement. If we accept the possibility that 'nothing' can be added or removed from s and still effect a change to s, the above formula is equivalent to two changes, applied in sequence, each expressed as the formula

$$s \rightarrow s - \mathrm{f}(a) + \mathrm{f}(b),$$

where a or b may refer to an 'empty' object.

We can stipulate conditions on the application of this formula. For instance, if we impose the condition that there is a connection between a and s—that is, there is a truth-functional ψ such that $s \rightarrow s - \mathrm{f}(a) + \mathrm{f}(b)$ can only be applied if $\psi[s, \mathrm{f}(a)]$ is satisfied—then we arrive at the familiar notion of a 'rule', $a \rightarrow b$. Typically, ψ is satisfied when certain aspects or properties of a 'occur' in s. If we further impose the condition that any and all changes are effected only through formulas of this form, the system then specifies a 'grammar'. Of particular importance is the notion of occurrence, for it is those *emergent* aspects or properties of the object of interest that one normally wishes to change. The derivation $s \rightarrow s - \mathrm{f}(a) + \mathrm{f}(b)$ may be viewed as representing a basic equation of spatial change, where $\mathrm{f}(a)$ is a representation of the emergent properties in s that is being altered.

We claim that this formula captures nearly every kind of spatial editing change. Let us consider, for example, that the change $s \rightarrow \mathrm{f}(s)$ specifies a simple transformation of s. Here the emergent shape is, simply, s. In practice we would replace s by $\mathrm{f}(s)$ directly rather than treat it as an application of the rule $s \rightarrow \mathrm{f}(s)$, under an identity transformation. The formula is interesting in two respects. First, it offers a simple mechanism for structuring change. Second, it specifies which aspects of s we have to look for in effecting the change. It is the latter that influences the way in which humans perceive objects and affects the way in which they perceive change. This perception is particularly important to designers in their explorations of spatial forms.

In this paper, the objects are shapes and the changes are spatial changes. Functions on shapes can be considered to be geometrical transformations. We can specify a *part relation*, whereby *any* part of a shape is a shape. That is, a shape identifies an indefinite number of shapes, each a part of the original shape. Shapes emerge under the part relation, even though these may not originally have been envisioned as such. Emergent shapes become explicit only when manipulated as such. Recognizing emergent shapes requires determining a transformation under which a shape is a part of the original shape.

The idea that designs are the products of evolving spatial changes has been considered by others—for example, by Bridges (1991) who informally explores the (presumed) role of computers in the design studio, and less informally by Oxman and Oxman (1991) who, again in a pedagogical context, advocate that an understanding of precedents can be achieved through analyses of spatial (transformational) change, which they refer to as refinement and adaptation. The notion of studying spatial change—either informally or formally—is not new. Figure 1 illustrates a design for a church by Leonardo da Vinci (Galluzzi, 1987). This is an example of many designs that Leonardo created with the irregular octagon, a form with which he was fascinated. It is a cross-shaped radiating plan with alternating chapels and niches. The satellite

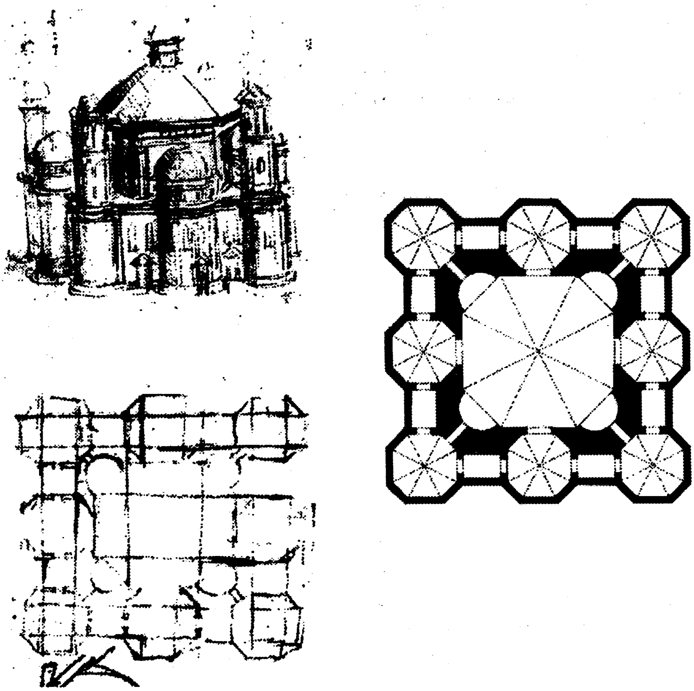


Figure 1. Design of a church by Leonardo da Vinci (circa 1500).

chapels connect diagonally converging into the central space. The octagon dominates the plan. It is interesting to see how Leonardo used the unevenly spaced grid and the (progressively added) axial lines to sculpt the emergent octagon (though he was not always successful in this endeavour). Figure 2 illustrates how an octagon emerges by the superposition (or addition) of the two sets of grid lines.

The plan illustrates the three aspects of spatial change that form the subject matter of this paper. First, the development of the plan through spatial transformation (change) of the underlying grid. Second, emergence of the octagons by delineating certain parts of the grid and axial lines. Third, the existence of an overall structure and continuity of composition.

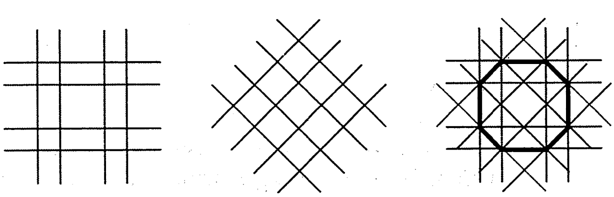


Figure 2. Emerging octagon by adding two sets of grid lines.

Shape rules and algebra

A shape rule is a mechanism that effects a spatial change. A shape rule $a \rightarrow b$ specifies a spatial relationship between a and b, which when applied to a shape s under a transformation f, such that $\mathrm{f}(a)$ is a part of s, replaces $\mathrm{f}(a)$ in s by $\mathrm{f}(b)$ under rule application. That is, when the shape rule is applied to the shape s, it produces the shape $s - \mathrm{f}(a) + \mathrm{f}(b)$. The set F of valid transformations is the set of all Euclidean transformations, which consist of translations, rotations, reflections, and scale.

A shape rule constitutes a formal specification of shape recognition and subsequent manipulation. For a shape rule $a \rightarrow b$, the left-hand side (a) specifies the similar shape to be recognized and the right-hand side (b) specifies the replacement leading to the resulting shape. A shape rule application consists of replacing the emergent shape corresponding to a, under some allowable transformation, by b, under the same transformation.

Shape rules operate on shapes within an algebra, U, which is closed under the operations of sum (shape union, '+'), difference ('−'), and product (shape intersection, '·'), and a set of transformations F. We define a part relation '$\leqslant$' on U such that $\mathrm{f}(a) \leqslant s$ whenever a is a shape in s for some member f of F.

Shapes form algebras under the part relation with well-defined properties (Stiny, 1991; Stouffs, 1994). We denote a shape algebra by U_{ij}, the set of all shapes made up of i-dimensional elements embedded in a j-dimensional Euclidean space E^j, $j \geqslant i$; by U_i if j is understood; and by U in general.

Shapes can be augmented in a number of ways, for instance, by distinguishing certain parts of the shape which introduce additional spatial relations. For example, we can attach labels from a given set to points. Labeled points formed in this way are ordered pairs that can be arranged into sets to specify, in a manner analogous to U_0, an algebra V_0 of labeled points. Spatial transformations of labeled points keep the labels the same though the points may alter. Likewise, labels attached to the same point combine under the sum operation. Thus, a *labeled shape* can be considered as an element of the algebra V, $V = U \times V_0$, which has the same properties as U. A labeled shape is made up of a shape and a finite, possibly empty, set of labeled points. Other augmentations of shapes are possible (Stiny, 1992).

Shape description and topology

When designers work with shapes they do so in a manner quite distinct from the (computational) representations for shapes. Often this has to do with the process of designing and with explanations for design choices or decisions. In other words, a shape is structured so as to provide a description for it. Typically this structuring takes the form of a decomposition of the shape into parts, the sum of which equals the shape.

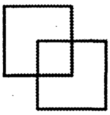
A particular form of decomposition of a shape corresponds to a *topology*, which is closed under sum and product (Hocking and Young, 1988). A topology for a shape satisfies two additional conditions. First, both the shape and the empty shape are in the topology. Second, for any shape x which is a part of the given shape, there is a smallest shape in the topology of which x is also a part. Every element of the topology is a part of the shape. Thus, the sum of the elements of the topology equals the shape. A topology for a shape specifies a way of cutting up a shape into a collection of fixed parts. A shape can have any number of topologies defined on it.

There are two obvious topologies for a shape s, the trivial topology consisting of the empty shape (0) and the shape (s), and the infinite topology made up of all parts of s. The more interesting typologies are those that cut up a shape somewhere in between these extremes (Stiny, 1994). Figure 3 illustrates a shape and a topology defined on it.

Related to topologies for a shape are closure relations defined on shapes. A *closure relation* c on a shape s is a mapping between parts of s that satisfies the following conditions:

(1) $\mathrm{c}(0) = 0$, and $\mathrm{c}(s) = s$;
(2) $x \leqslant \mathrm{c}(x)$;
(3) $\mathrm{c}[\mathrm{c}(x)] = \mathrm{c}(x)$;
(4) $\mathrm{c}(x+y) = \mathrm{c}(x) + (y)$.

shape



topology

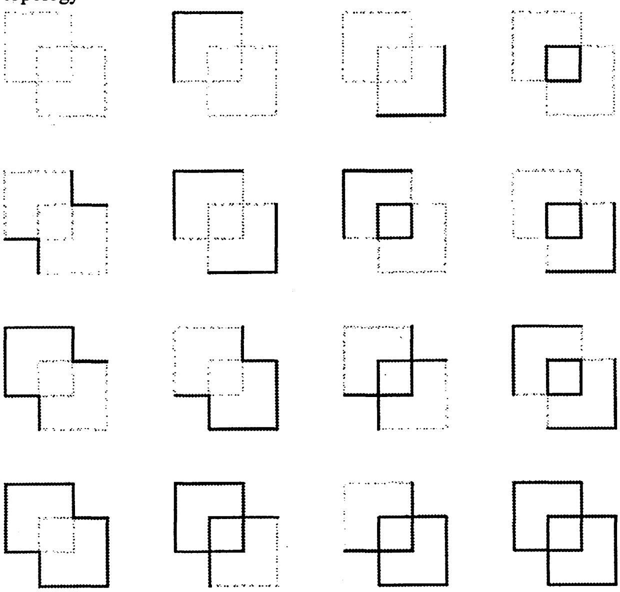


Figure 3. A shape and a possible topology for it.

In addition, a closure relation c has the following properties.
(5) $x \leqslant y \Rightarrow \mathrm{c}(x) \leqslant \mathrm{c}(y)$;
(6) $\mathrm{c}(x \cdot y) \leqslant \mathrm{c}(x) \cdot \mathrm{c}(y)$.

We denote the topology for a shape s by T_s, and its closure relation by c_s. The connection between a topology and its closure relation is strong in the sense that each specifies the other. For any part x, $x \leqslant s$, its closure $\mathrm{c}_s(x)$ is an element of T_s. In fact, every element of T_s is a *closed shape.* For shape s, with closure relation c_s,

$$T_s = \bigcup_{x \leqslant s} \mathrm{c}_s(x) .$$

New shapes can be produced from a shape by the application of shape rules. A shape rule may be considered as a mapping between two shapes. Alternatively, as shown below, a shape rule relates the descriptions (topologies) of the two shapes. A computation, from one shape to another, is a series of shape rule applications that results in the production of the second shape from the first.

Consider a shape s with topology T_s. Suppose the shape rule $a \rightarrow b$ is applicable to s. Then there is a transformation f such that $\mathrm{f}(a) \leqslant s$. Under shape rule application, we have

$$t = s - \mathrm{f}(a) + \mathrm{f}(b) .$$

There are two distinct ways in which such rule applications can be considered: constructively and apperceptively.[2]

[2] 'Doing' and 'seeing', as George Stiny puts it (personal communication).

It is perhaps instructive first to examine the constructive effect of spatial change that is wrought by rule application. We treat a rule as a function that 'acts' on parts of s. There are a number of ways in which such a function can be expressed. We note that t can be rewritten as:

$$t = s - \mathrm{f}(a-b) + \mathrm{f}(b-a) .$$

That is, $a \rightarrow b$ has the same effect as the rule $(a-b) \rightarrow (b-a)$, where the left-hand and right-hand shapes are now disjoint. However, $\mathrm{f}(b-a)$ may have parts in common with s. Thus,

$$t = s - \mathrm{f}(a-b) + [\mathrm{f}(b-a) - s] .$$

Here, $\mathrm{f}(a-b)$ is a part of s, $[\mathrm{f}(b-a)-s]$ has no part in common with s, and the rule is reduced to one in which the left-hand and right-hand shapes are disjoint and the replacement shape has no part in common with the given shape.

We can characterize this behaviour of rule application by the function:

$$\mathrm{h}(x) = \begin{cases} x - \hat{a} + \hat{b}, & \text{if } x \cdot \hat{a} \neq 0, \\ x, & \text{otherwise}, \end{cases} \tag{1}$$

where $\hat{a} = \mathrm{f}(a-b)$, and $\hat{b} = \mathrm{f}(b-a) - s \equiv \mathrm{f}(b) - s$. Note that h affects those parts of s that have something in common with $\mathrm{f}(a)$ but not in common with $\mathrm{f}(b)$. The rule $\hat{a} \rightarrow \hat{b}$ provides a representation of the shape rule $a \rightarrow b$ under the transformation f, $\hat{a}$ is a representation of the emergent shape, $\hat{b}$ is a representation of the replacement shape, and h is a representation of the application of the shape rule. The properties of h are given in table 1. The proof follows directly from the definition of h.

Table 1. Properties of h.

$x = 0$	$\mathrm{h}(0) = 0$
$x = s$	$\mathrm{h}(s) = t$
	$\mathrm{h}(x+y) = \mathrm{h}(x) + \mathrm{h}(y)$
$x, y \leqslant s$	$\mathrm{h}(x \cdot y) \leqslant \mathrm{h}(x) \cdot \mathrm{h}(y)$
	$x \leqslant y \Rightarrow \mathrm{h}(x) \leqslant \mathrm{h}(y)$

It is important to note that, although $\hat{a} \rightarrow \hat{b}$ has the same effect as $a \rightarrow b$, these are not equivalent rules either constructively and apperceptively. The rule $\hat{a} \rightarrow \hat{b}$ applies to more spatial situations in s than does the rule $a \rightarrow b$. Equally, it is important to note that $\hat{a} \rightarrow \hat{b}$ is its own representation under the transformation f.

The function h maps parts of s to parts of t, and in the process may be considered to induce a division of t into a set of parts, T_t. For the closure relation c_s on s, we can consider the relation c_t:

$$\mathrm{c}_t(x) = \begin{cases} \mathrm{h}[\mathrm{c}_s(x \cdot s)] + \hat{b}, & \text{if } x \cdot \hat{b} \neq 0, \\ \mathrm{h}[\mathrm{c}_s(x)], & \text{otherwise.} \end{cases}$$

Note that, if $x \cdot \hat{b} = 0$, $x \leqslant s$. The relation c_t basically leaves untouched those closed shapes of s which have no part in common with the replacement shape; otherwise, it adds $\hat{b}$ to each closed shape. Relation c_t satisfies the properties of a closure relation on t as proven below.

We note that $\mathrm{c}_t(0) = \mathrm{h}[\mathrm{c}_s(0)] = \mathrm{h}(0) = 0$. We have

$$\mathrm{c}_t(t) = \mathrm{h}[\mathrm{c}_s(x \cdot s)] + \hat{b} \leqslant \mathrm{h}[\mathrm{c}_s(s)] + \hat{b} = \mathrm{h}(s) + \hat{b} = t .$$

That is, $\mathrm{c}_t(t) \leqslant t$.

If $(x+y)\cdot\hat{b} = 0$, then

$$c_t(x+y) = h[c_s(x+y)] = h[c_s(x)+c_s(y)] = h[c_s(x)]+h[c_s(y)] = c_t(x)+c_t(y).$$

Otherwise

$$c_t(x+y) = h\{c_s[(x+y)\cdot s]\}+\hat{b} = h[c_s(x\cdot s)]+h[c_s(y\cdot s)]+\hat{b} = c_t(x)+c_t(y).$$

We now show that, for $x \leqslant t$, $x \leqslant c_t(x)$. We note that $x\cdot\hat{a} = 0$. If $x\cdot\hat{b} = 0$, $h(x) = x$; $x \leqslant c_s(x) \Rightarrow h(x) \leqslant h[c_s(x)] \Rightarrow x \leqslant c_t(x)$. If $x \leqslant \hat{b}$, then

$$x \leqslant h[c_s(x\cdot s)]+\hat{b} = c_t(x).$$

If neither holds, then x can be partitioned with respect to $\hat{b}$, $x = (x-\hat{b})+(x\cdot\hat{b})$, and the inequality naturally follows.

From $c_t(t) \leqslant t$ and $t \leqslant c_t(t)$, we get $c_t(t) = t$, and $c_t(\hat{b}) = h[c_s(\hat{b}\cdot s)]+\hat{b} = \hat{b}$. We have only $c_t[c_t(x)] \leqslant c_t(x)$ left to consider. Here we have two cases. Suppose $x\cdot\hat{b} = 0$, and $c_t(x) = h[c_s(x)]$. If $c_s(x)\cdot\hat{a} = 0$, then

$$c_t(x) = c_s(x),$$
$$c_t[c_t(x)] = c_t[c_s(x)] = h\{c_s[c_s(x)]\} = h[c_s(x)] = c_t(x).$$

Otherwise

$$c_t(x) = c_s(x)-\hat{a}+\hat{b},$$
$$c_t[c_t(x)] = c_t[c_s(x)-\hat{a}]+c_t(\hat{b}) = h\{c_s[c_s(x)-\hat{a}]\}+\hat{b} \leqslant h[c_s(x)]+\hat{b}$$
$$= c_s(x)-\hat{a}+\hat{b} = c_t(x).$$

If $x\cdot\hat{b} \neq 0$, then

$$c_t(x) = h[c_s(x\cdot s)]+\hat{b} = c_s(x\cdot s)-\hat{a}+\hat{b},$$
$$c_t[c_t(x)] = c_t[c_s(x\cdot s)-\hat{a}]+c_t(\hat{b}) = h\{c_s[c_s(x\cdot s)-\hat{a}]\}+\hat{b} \leqslant h[c_s(x\cdot s)]+\hat{b}$$
$$= c_s(x\cdot s)-\hat{a}+\hat{b} = c_t(x).$$

Whence, the set $T_t = \bigcup_{x \leqslant t} c_t(x)$ is a topology for shape t.

Rule application and continuity

Stiny (1994) offers an interesting characterization of shape rule application that is related to the way a shape is cut up into parts to form a description. In his paper he gives a handsomely deconstructivist interpretation whereby the present (shape) explains the precedent (shape) which justifies the present, and does so through the concept of continuity. We provide further elaboration.

Two topologies are related by a mapping that is *continuous* whenever each closed shape in one topology is mapped into a closed shape in the other (Stiny, 1994). We consider the mapping h that relates two shapes s and t, and hence their respective topologies T_s and T_t. Formally, h is *continuous* whenever, for $x \leqslant s$, $h[c_s(x)] \leqslant c_t[h(x)]$, where c_s and c_t are the closure relations in T_s and T_t. The continuity of h is now explored.

Let us consider any part $x \leqslant s$. Let $x\cdot\hat{b} = 0$. If $x\cdot\hat{a} = 0$, then $h(x) = x$, $c_t[h(x)] = c_t(x) = h[c_s(x)]$. Otherwise, consider a partitioning of x with respect to $\hat{a}$: $x = x-\hat{a}+x\cdot\hat{a}$. Then

$$h(x) = x-\hat{a}+\hat{b},$$
$$c_t[h(x)] = c_t(x-\hat{a})+c_t(\hat{b}) = h[c_s(x-\hat{a})]+h[c_s(\hat{b})],$$
$$h[c_s(x)] = h[c_s(x-\hat{a})+c_s(x\cdot\hat{a})] = h[c_s(x-\hat{a})]+h[c_s(x\cdot\hat{a})].$$

Whence

$$\begin{aligned} & \mathrm{h}[\mathrm{c}_s(x)] \leqslant \mathrm{c}_t[\mathrm{h}(x)] \\ & \quad \Leftrightarrow \mathrm{h}[\mathrm{c}_s(x-\hat{a})]+\mathrm{h}[\mathrm{c}_s(x\cdot\hat{a})] \leqslant \mathrm{h}[\mathrm{c}_s(x-\hat{a})]+\mathrm{h}[\mathrm{c}_s(\hat{b})] \\ & \quad \Leftrightarrow \mathrm{h}[\mathrm{c}_s(x\cdot\hat{a})] \leqslant \mathrm{h}[\mathrm{c}_s(x-\hat{a})]+\mathrm{h}[\mathrm{c}_s(\hat{b})] \\ & \quad \Leftrightarrow \mathrm{c}_s(x\cdot\hat{a})-\hat{a}+\hat{b} \leqslant \mathrm{c}_s(x-\hat{a})-\hat{a}+\hat{b} \\ & \quad \Leftrightarrow \mathrm{c}_s(x\cdot\hat{a}) \leqslant \mathrm{c}_s(x-\hat{a})+\hat{a}, \qquad \text{because } \mathrm{c}_s(x\cdot\hat{a})\cdot\hat{b}=0. \end{aligned}$$

As this must hold for any $x \leqslant s$, it must also hold for $x = \hat{a} \Rightarrow \mathrm{c}_s(\hat{a}) = \hat{a}$. Furthermore, for $x \leqslant s$,

$$\mathrm{c}_s(x\cdot\hat{a}) \leqslant \mathrm{c}_s(\hat{a}) \quad 0 \leqslant \mathrm{c}_s(x-\hat{a}),$$

and thus

$$\mathrm{c}_s(x\cdot\hat{a}) \leqslant \mathrm{c}_s(\hat{a}) = \hat{a} \leqslant \mathrm{c}_s(x-\hat{a})+\hat{a}.$$

It follows that $\mathrm{c}_s(\hat{a}) = \hat{a}$ is a necessary and sufficient condition for the rule application to be continuous. That is, the representation of the emergent shape has to be closed in the topology of s. However, this condition is dependent on the specific choice of the function h. For different choices of h, different conditions exist for the rule application to be continuous. For instance, if we choose h as

$$\mathrm{h}(x) = \begin{cases} \{x+\mathrm{c}_s[\mathrm{f}(b)\cdot s]\}-\mathrm{f}(a)+\mathrm{f}(b), & \text{if } x\cdot\mathrm{f}(a) \neq 0, \\ x, & \text{otherwise,} \end{cases} \tag{2}$$

and c_t as

$$\mathrm{c}_t(x) = \begin{cases} \mathrm{h}\{\mathrm{c}_s(x\cdot s)+\mathrm{c}_s[\mathrm{f}(b)\cdot s]\}+\mathrm{f}(b), & \text{if } x\cdot[\mathrm{f}(b)-s] \neq 0, \\ \mathrm{h}[\mathrm{c}_s(x)], & \text{otherwise,} \end{cases}$$

we can show that h [as defined in equation (2)] satisfies the properties in table 1, and that c_t specifies the closure relation for a topology for t. It is important to note that the term $\mathrm{c}_s[\mathrm{f}(b)\cdot s]$ is required to account for all parts of the replacement shape $\mathrm{f}(b)$ that has parts in common with s. In this case we can show, following an argument similar to the one above, that $\mathrm{c}_s[\mathrm{f}(a)] \leqslant \mathrm{f}(a)+\mathrm{c}_s[\mathrm{f}(b)\cdot s]$ is a necessary and sufficient condition for the rule application to be continuous. Here, the condition for continuity of shape rule application is independent of x but dependent on s, $\mathrm{f}(a)$, and $\mathrm{f}(b)$, as well as the topology defined on s by the closure relation c_s. There are two cases to consider with respect to this condition.

There are a set of rules for which rule application is always continuous, regardless of the topology defined. These are strictly additive rules for which $a \leqslant b$, or $\mathrm{c}_s[\mathrm{f}(a)] \leqslant \mathrm{c}_s[\mathrm{f}(b)\cdot s]$. Note that this is equivalent to the condition $\hat{a} = 0$. Any rule application may be considered continuous if we choose the topology for s carefully so that $\mathrm{c}_s[\mathrm{f}(a)] \leqslant \mathrm{c}_s[\mathrm{f}(b)\cdot s]+\mathrm{f}(a)$. We can examine this condition more closely for its effect on a topology.

Suppose we partition $\mathrm{f}(a)$ with respect to $\mathrm{c}_s[\mathrm{f}(b)\cdot s]$. We then have

$$\mathrm{c}_s[\mathrm{f}(a)] = \mathrm{c}_s\{\mathrm{f}(a)-\mathrm{c}_s[\mathrm{f}(b)\cdot s]\}+\mathrm{c}_s\{\mathrm{f}(a)\cdot\mathrm{c}_s[\mathrm{f}(b)\cdot s]\}.$$

The second term is always a part of $\mathrm{c}_s[\mathrm{f}(b)\cdot s]$; thus, we can rewrite the condition as

$$\mathrm{c}_s\{\mathrm{f}(a)-\mathrm{c}_s[\mathrm{f}(b)\cdot s]\} \leqslant \mathrm{f}(a)+\mathrm{c}_s[\mathrm{f}(b)\cdot s].$$

Given the properties of the closure operation, we can restate this condition upon the elements of the topology as follows: there exists a closed shape y of the topology T_s, such that

$$f(a) - c_s[f(b) \cdot s] \leqslant y \leqslant f(a) + c_s[f(b) \cdot s] .$$

If $a \leqslant b$ then $f(a) - c_s[f(b) \cdot s] = 0$, and $y = 0$ becomes a solution. As the empty shape 0 is always an element in the topology T_s, additive rules are always continuous, as shown before.

We note that $y = f(a-b) = \hat{a}$ constitutes a solution.

If $f(b) \cdot s = 0$, there is exactly one solution, namely, $y = f(a)$. Note that this is similar to the situation considered previously; here, $\hat{a} = f(a)$ and $\hat{b} = f(b)$. Table 2 gives the continuity conditions on h as defined in equation (2). In general, the specific range of solutions depends on the closure relation, as we need to evaluate $c_s[f(b) \cdot s]$.

Table 2. Continuity of h.

$$h(x) = \begin{cases} \{x + c_s[f(b) \cdot s]\} - f(a) + f(b) & \text{if } x \cdot f(a) \neq 0, \\ x, & \text{otherwise} \end{cases}$$

$a \leqslant b$

$$\left.\begin{array}{l} f(a) \in T_s \\ f(a-b) \in T_s \\ f(a+b) \cdot s \in T_s \end{array}\right\} \begin{array}{l} y \in T_s \\ f(a) - c_s[f(b) \cdot s] \leqslant y \leqslant f(a) + c_s[f(b) \cdot s] \end{array}$$

Discussion

Stiny demonstrates the relationship between continuity and shape emergence in the following way. Every shape rule computation can be made continuous provided one can structure topologies (and hence descriptions) for shapes so that emergent shapes are distinguished in the descriptions. He offers a construction in which every topology in the sequence representing a series of shape rule applications contains as an element the emergent shape that is altered by the corresponding shape rule in the sequence. It should be noted that the sequence of topologies is induced retroactively.

Stiny develops his formulation in order to give an account of emergence in shape computation. As a consequence, it is necessary that the emergent shapes are distinguished as such in descriptions of shapes. For a shape rule $a \rightarrow b$, $f(a)$ is generally considered the emergent shape but any shape that is a part of $f(a+b) \cdot s$, and has $f(a-b)$ as a part, can be considered a representation of the emergent shape. Thus, *for a computation to be continuous, a representation of the emergent shape has to be closed in each topology in the series that defines the computation.* As the analysis above indicates, this requirement is strong. In other words, continuity of computation requires anticipation of the emergent shapes that are to be changed.

This proposition complements Stiny's original result. Although every shape rule computation can be made continuous, retroactive induction illustrates only the potential for continuity; for computations to be continuous, representations of the emergent shapes have to be closed (and thus anticipated either way) in the descriptions within a computation.

It is important to note that the solution $f(a) \in T_s$ is independent of b but it is not independent of c_s, namely, the way in which a shape is decomposed into its description. The solution $f(a) \in T_s$ is a continuity condition for all shape rules in the form $a \rightarrow k$. In fact, in this case it is the only solution. Note that this condition does not presuppose any conditions on the replacement shape and s.

The solution $f(a) \in T_s$ is the most interesting and unarguably the most intuitive. It is also the condition that best draws out the apperceptive nature of shape rule application. Consider the spatial change under an application of the rule $a \rightarrow a$. In fact s does not change. Yet sequences of spatial changes that include rules of the form $a \rightarrow a$ will not be continuous unless $f(a) \in T_s$. That is, in order to ensure continuity of computation, in general, descriptions of s have to be altered even when no spatial change has occurred. The shape rule $a \rightarrow a$ may be viewed as reflecting the situation where designers pause and 'contemplate' the design thus far. The mere act of observing one's design might have the effect of changing its description. This remark—as it pertains to the act of designing—enters the realms of philosophy and cognitive psychology, areas in which we would not consider ourselves competent (for example, see Stiny, 1996).

One can of course impose an arbitrary topology on shapes where the emergent shape is always an element and restrict shape rule application to known elements. This then ensures that, for each such 'object-oriented' view of a shape, computations involving known objects (within that view) will be continuous. However, these will also be independent of precedent. If, as Stiny shows, one wants 'interesting' computations that distinguish emergent shapes, then these, by necessity, involve precedent and are only object-oriented after the fact.[3]

The above result is not altogether surprising for spatial computations involving object-oriented descriptions. Taking any standard textbook definition of objects, it is straightforward to identify an isomorphism between objects and an augmented algebra of shapes made up of 'point' figures, $A_0 \subseteq U_0 \times A$, where the elements of U_0 are indices to the objects. Set A may of course reflect spatial and nonspatial attributes. Computation is essentially defined on finite sets of points. Technical difficulties, if any, lie in the manner in which elements of A can be operated on, for example, the inheritance of attributes or properties. For spatial computations to be interesting, notions of emergence have to be introduced into the definition and treatment of objects.

It is possible to ensure continuous computations by always distinguishing the emergent shape $f(a)$. In certain situations this can be done by restructuring the topologies prior to each shape rule application, and always if this restructuring is carried out retroactively following Stiny's procedure.

The restructuring of descriptions opens up interesting issues in design thinking. For instance, if we accept the hypothesis that descriptions convey structure and meaning—in other words, the topology sits within a system of features such as hierarchical classification schemes, for example, semantic networks or objected-oriented systems

[3] The inherent dynamism of descriptions is a property that does not belong solely to the realm of design. There are striking parallels in other areas that demonstrate the connection between continuity, precedent, and emergence. Stiny cites sources in law where one seeks articulate consistency or, at the very least, continuity from precedent in order to explain a current legal position. In genetics, hereditary diseases such as cystic fibrosis or sickle-cell anaemia have been linked to certain DNA patterns within genes. Researchers have attempted to establish historical precedents for present-day patterns that can be explained through evolutionary gene mutation—in doing so, they have sometimes reclassified or redefined their taxonomies. The historian Edward Hallett Carr makes the same point in the traditional field of historical research. In his book, *What is History* Carr (1961, pages 35, 164) declares history to be "... a *continuous* process of interaction between the historian and his facts, an unending ... dialogue between the events of the past and the progressively emerging future ends. The historian's interpretation of the past, his *selection* of significant and the relevant, evolves with the progressive *emergence* of new goals" [our italics]. It seems that valid classifications, which can 'explain' designs, are equally postrationalised.

made up of known components with known semantics—and given that all spatial changes can be effected by mechanisms of the form given by $s \rightarrow s - \mathrm{f}(a) + \mathrm{f}(b)$, the preceding analysis suggests that there is a distinction to be drawn between descriptions that designers use while designing and descriptions that they employ to explain their designs. This may explain the 'discrepancy' which often surfaces between the avowed process, as evidenced by their stated descriptions and reasons, that designers claim to adopt at the start of their design and the actual process, again as evidenced by their stated descriptions and reasons, that they follow in arriving at their designs. If we accept that continuity of descriptions is a measure of articulate consistency, then this may explain that the reason why object-oriented approaches and case-based reasoning are becoming increasingly popular in CAD is precisely because operations on such descriptions are continuous. However, the analysis shows that these systems will be bereft of novelty, where novelty arises in situations in which the designer does not view a spatial entity as a fixed object with fixed descriptions but perceives it to be malleable and it can thus be reshaped to produce new spatial (and, consequently, new semantic) relationships. It is these emergent shapes made up of parts from known objects that, we believe, contribute to novelty in design.

Reversible and irreversible rules

Rules combine to form *grammars* which are formal rewriting systems for producing objects of interest. A shape grammar combines a set of shape rules and defines a *language*, which is the set containing all shapes, generated by the grammar, that have no associated symbols. Figure 4 shows a few shapes produced from an initial shape by just a single rule. All shapes produced are members of the language of the corresponding grammar.

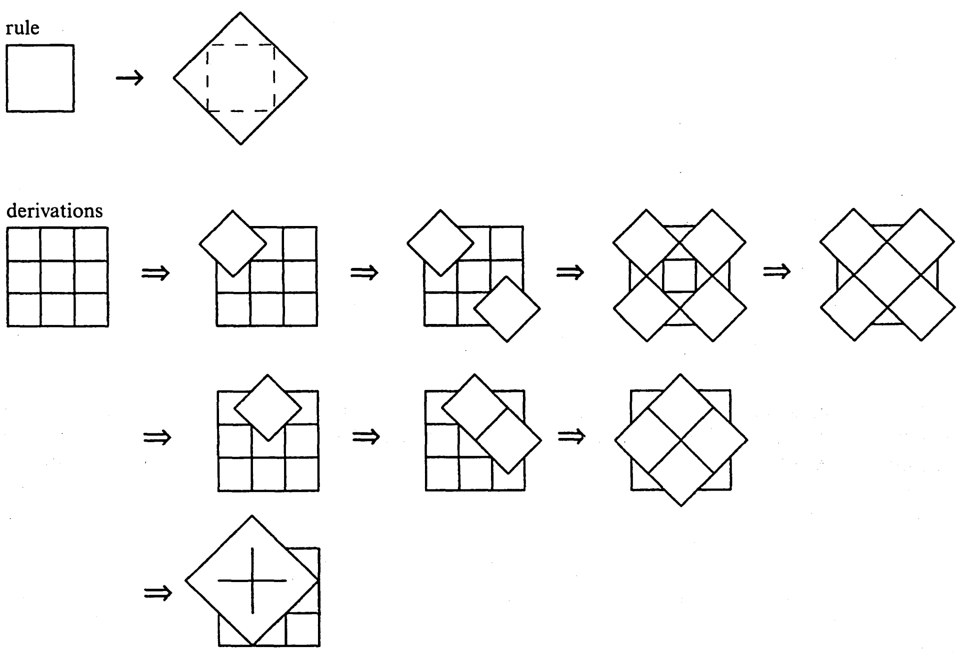


Figure 4. Exemplar derivations from an initial shape with a single rule.

Computationally, an important issue is whether one can easily backtrack along a computation to a shape from which alternative spatial forms may be explored. Related to this is the following question: whether for any rule in a grammar a reverse rule can be constructed such that, when the original and reverse rules are applied, consecutively, to any element of the algebra over which the grammar is defined, the result is identical to the original element. The rules are all reversible, the corresponding grammar is said to be reversible.

The ability to reverse spatial computations is important for systems in which one can explore design space. If computations are reversible, then a simple recording of the changes that have been hitherto invoked is all that is required to return to any previous state. If computations are not reversible, additional shape information has to be recorded. The question then is whether the price of additional bookkeeping may be offset by the greater flexibility that irreversible rules offer for exploring the world of possible designs.

We now consider the reversibility of shape rule application. Consider a shape rule $a \rightarrow b$ that applies to a shape s. Then

$$t = s - \mathrm{f}(a) + \mathrm{f}(b) = s - \mathrm{f}(a-b) + \mathrm{f}(b-a) = s - \mathrm{f}(a-b) + [\mathrm{f}(b-a) - s] \,.$$

Assume that there is a rule $x \rightarrow y$, which may be identical to $b \rightarrow a$, such that the shape that results from applying the rules $a \rightarrow b$ and $x \rightarrow y$ to s, in that order and under the same transformation f, equals s. We can assume that the rule $x \rightarrow y$ applies under the same transformation; otherwise we can always transform the rule such that it applies under the same transformation, without changing the rule application or its scope. Thus

$$s = t - \mathrm{f}(x) + \mathrm{f}(y) = t - \mathrm{f}(x-y) + \mathrm{f}(y-x) = t - \mathrm{f}(x-y) + [\mathrm{f}(y-x) - t] \,.$$

Whence

$$\begin{aligned} s &= \{s - \mathrm{f}(a-b) + [\mathrm{f}(b-a) - s]\} - \mathrm{f}(x-y) + [\mathrm{f}(y-x) - t] \\ &= s - \mathrm{f}(a-b) - \mathrm{f}(x-y) + [\mathrm{f}(b-a) - s - \mathrm{f}(x-y)] + [\mathrm{f}(y-x) - t] \,. \end{aligned}$$

As $\mathrm{f}(x-y) \cdot s = 0$, and $[\mathrm{f}(b-a) - s] \cdot s = 0$, the above equation implies the following:

$$\mathrm{f}(b-a) - s - \mathrm{f}(x-y) = 0, \qquad s = s - \mathrm{f}(a-b) + [\mathrm{f}(y-x) - t] \,.$$

Thus

$$\mathrm{f}(b-a) - s \leqslant \mathrm{f}(x-y), \qquad \mathrm{f}(a-b) \leqslant \mathrm{f}(y-x) - t \,.$$

Similarly, substituting the expression for s in the expression for t, we obtain:

$$\mathrm{f}(x-y) \leqslant \mathrm{f}(b-a) - s, \qquad \mathrm{f}(y-x) - t \leqslant \mathrm{f}(a-b) \,.$$

Therefore

$$\mathrm{f}(x-y) = \mathrm{f}(b-a) - s, \qquad \mathrm{f}(y-x) - t = \mathrm{f}(a-b) \,.$$

These two equations give a specification for x and y in terms of the rule $a \rightarrow b$, as well as the shape s under application. The rule $x \rightarrow y$ is independent of s only if $\mathrm{f}(b-a) \cdot s = 0$, so that $\mathrm{f}(x-y) = \mathrm{f}(b-a)$. Similarly, note that $a \rightarrow b$ is the reverse rule for $x \rightarrow y$ for all shapes t only when $\mathrm{f}(y-x) \cdot t = 0$. This condition was first given in Krishnamurti (1981) though not formally proved there.

Discussion

Irreversibility of shape rules distinguishes shape grammars from most other grammar formalisms. Intuitively we note that, when two shapes (or sets) are combined under the operation of $+$, identical elements 'merge'. That is, only a single occurrence of the element appears in the resulting shape or set. In the case of set grammars, if $|u|$ denotes the cardinality of a set u,

$$\text{maximum}(|u|,|v|) \leqslant |u+v| \leqslant |u|+|v| .$$

On the other hand, in the case of string or graph grammars (given an appropriate definition for the size of a graph) this would constitute strict equality. No comparable measure exists for shapes except for shapes defined in U_0.

The condition $\mathrm{f}(b-a)\cdot s = 0$ is an expression of this situation. The transformation $\mathrm{f}(b-a)$ denotes the shape that is added to s, which is not previously removed, under rule application. If the product with s equals zero, no elements merge and the rule application is reversible. The condition is in general dependent on the particular rule application, that is, on the original shape s and transformation f. When $b \leqslant a$, $\mathrm{f}(b-a) = 0$, and thus the rule is reversible independent of the shape to which (and the transformation under which) it is applied. These constitute purely subtractive rules. It is interesting to note that if $a \rightarrow b$ is a reversible rule its reverse rule $b \rightarrow a$ is not unless $b = a$.

Figures 5 and 6 (see over) illustrate a reversible and an irreversible shape rule, respectively. In general, shape rules are irreversible. Yet every shape rule computation can be reversed provided one records the *mergent* shape, $\mathrm{f}(b-a)\cdot s$, for each rule application in the computation such that these shapes can be added, appropriately, in a reversal of computation. In other words—unlike in most spatial computational systems—it does not suffice merely to record the acts of change, but it is also necessary to record a representation of the spatial changes along with the acts. Thus, like continuity, reversibility of computation requires anticipation. In this case it requires an anticipation of the mergent shape, $\mathrm{f}(b-a)\cdot s$, within the original shape, s.

The significance from a design standpoint—at any rate, to us—stems from the fact that designing is a temporal activity. The irreversibility of a rule has the effect of time stamping each rule application and thus capturing design 'intent' at any given time.

It should be noted that, whenever a reversible rule $a \rightarrow b$ is applied in a computation, shape b does not add to the description (topology) of the resulting shape. It should be further noted that the application of the reverse rule $b \rightarrow a$ at a

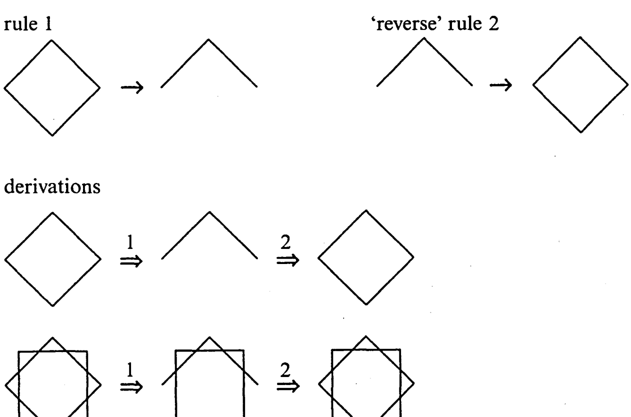


Figure 5. An example of a reversible rule $a \rightarrow b$ (rule 1). We observe from the exemplar derivations that, when the rules $a \rightarrow b$ and $b \rightarrow a$ (rule 2) are applied subsequently and under the same transformation, the resulting shape equals the original shape.

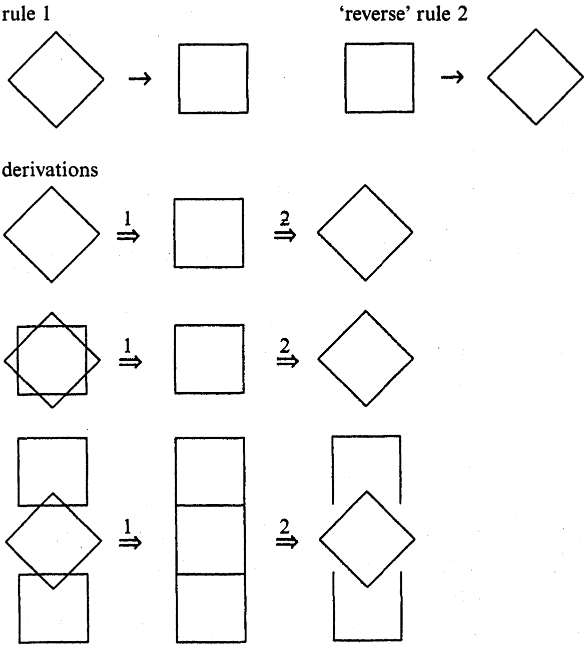


Figure 6. An example of an irreversible rule $a \rightarrow b$ (rule 1). We observe from the exemplar derivations that, when the rules $a \rightarrow b$ and $b \rightarrow a$ (rule 2) are applied subsequently and under the same transformation, the resulting shape may not equal the original shape.

subsequent stage in the computation under the same transformation f removes any trace of the effect of the original application on the design. However, in the case of an irreversible rule $a \rightarrow b$, the application of the rule $b \rightarrow a$ under the same transformation f does not remove the effect of the original rule application on the design. Moreover, for reversible rules, the computation to produce a given design may be indifferent to the ordering of rule applications. This is because the parts of the shape that are affected by the rules remain the same. This is not the case with irreversible rules. Thus, general shape rule application is both spatial and temporal.

Shape recognition

We have thus far examined two distinct aspects of the application of shape rule $a \rightarrow b$ to shape s. In our analyses we have assumed the existence of an affine transformation f such that $\mathrm{f}(a) \leqslant s$. We now turn our attention to the question of determining all transformations that satisfy the subshape relation, $\mathrm{f}(a) \leqslant s$. In general, *shape recognition* relates to finding one or all valid transformations under which a shape is a part of a given shape. In the case of shape rule application, the solution to this problem consists of finding a correspondence between the spatial elements on the left-hand side of the rule (a) and elements of the given shape (s), and determining the transformation f that represents this correspondence.[4] This is a difficult problem because a shape, with definite description, has an indefinite number of 'touchable'

[4] Some readers may be familiar with the equivalent problem in set or graph grammars, respectively termed *subset* and *subgraph detection*, that consists of searching for either a single entity or a group of entities within a set or a graph. Such a search is straightforward; it requires a one-to-one matching of entities that are identical under a certain transformation. The determination of the matching is not necessarily efficient, for example, subgraph isomorphism is NP-complete. On the other hand, a prerequisite for shape rule computation is that *any* subshape of a shape is spatially replaceable.

Table 3. Table of shape algebras.

U_0	U_{00},	U_{01},	U_{02},	U_{03},	...
U_1		U_{11},	U_{12},	U_{13},	...
U_2			U_{22},	U_{23},	...
U_3				U_{33},	...

parts. In respect to this problem, we take the position that shapes are *individuals* as reflected by the part relation defined on shapes.(5)

We consider shape recognition in each shape algebra and across shape algebras (table 3). Shape recognition in U_0 is equivalent to set recognition and is thus trivial. For shapes in U_1, *distinguishable points* serve as the basis for reducing the problem (Chase, 1989; Krishnamurti, 1981; Krishnamurti and Earl, 1992). For any given shape, its distinguishable points correspond to points (labeled or otherwise) the properties of which (with respect to the shape) are preserved under the part relation and affine transformations. Points of intersection of segments in a shape are distinguished. The application of this concept to U_i, for $i > 0$, is explored below.

In U_{ii} all shapes are necessarily coequal. Consequently, no distinguishable points, lines, or planes can be constructed. As a result, an indeterminate number of transformations exist for $i > 0$, the base transformation of which is the identity transformation. In U_{00} only the identity transformations exist.

In j dimensions, a correspondence between $j + 1$ not cohyperplanar distinguishable points uniquely determines an affine transformation. However, to determine all valid similarity transformations, j such points suffice provided the corresponding 'point figures' are similar to each other. Each such transformation remains subject to evaluation with respect to the shape a. Otherwise, if j cohyperplanar distinguishable points cannot be determined then there may be an indeterminate number of valid transformations. In such cases, one can isolate a set of base transformations from which the other transformations can be generated. Figure 7 illustrates this indeterminacy of transformations for shapes in U_{23} (see case 2.2 below).

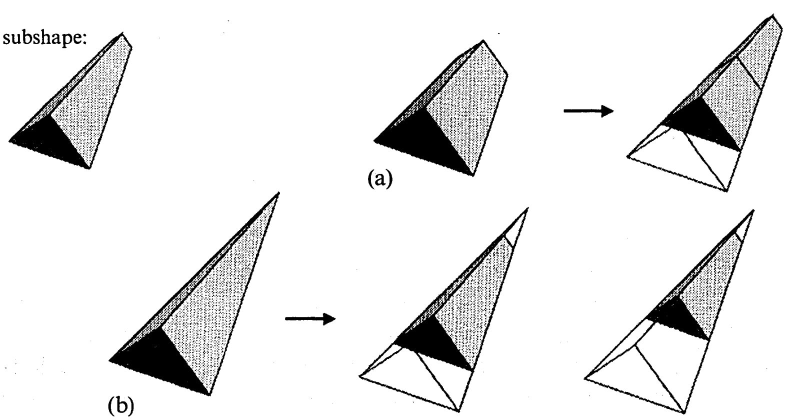


Figure 7. The base transformation maps carriers onto carriers without scaling: (a) no possible transformation exists, even under scaling; (b) an infinite number of possible transformations exist under scaling.

(5) The concept of individuals differs from the generally accepted concept of classes (or sets) in that no subdivision into subclasses or members is established or suggested a priori (Leonard and Goodman, 1940). Stiny (1993b), in a reversal of his earlier position (Stiny, 1982), offers a comparison between shapes and individuals, which differ algebraically. However, shapes and individuals can both be divided into parts in any way whatsoever.

In the sequel we restrict discussion to similarity transformations. Note that, even in the case of shapes with rational descriptions, the resulting transformations may not be rational (Krishnamurti and Earl, 1992). As such, the constructions described below do not guarantee exact arithmetic.

Detecting emergent shapes in U_{i3}

We give the determinate and indeterminate cases for shapes defined on U_{i3}, $i < 3$. As the representation of a shape in U_{i3} is embedded in E^3, we distinguish the case in E^3 by using as *primary* distinguishable elements the distinct carriers of the segments in the shape to be recognized. Additional distinguishable points may be constructed for the purpose of generating a determinate number of valid transformations, though the cases themselves are identified by the primary distinguishable elements. The cases are grouped by their shape algebra and numbered accordingly: the first digit denotes the dimensionality of the algebra. In the sequel, a denotes the shape to be recognized in the given shape s.

The cases for U_{03}

The cases are trivial. There is one determinate case when at least three noncolinear (distinguishable) points in the given shapes can be found. Otherwise, a possible indeterminate number of valid transformations between the shapes exist.

Case 0.1: There are three noncolinear points. Three noncolinear points uniquely determine an affine transformation in E^2, that is, for the plane these define. In E^3, if the point figures are similar this results in two Euclidean transformations, of which one can be derived from the other by use of a reflection in this plane. A fourth, noncoplanar, point suffices to invalidate one of these transformations. The computation of both transformations is described in Krishnamurti and Earl (1992).

Case 0.2: There are two (distinct) points. An indeterminate number of valid transformations exist. Given a base transformation that maps both points in the shape to be recognized onto the respective points in the given shape, the full set of possible transformations is derived from rotations about the line connecting both points and a reflection in a plane through this line.

Case 0.3: There is a single point. An indeterminate number of valid transformations exist. These can be derived from a base transformation by means of (three-axes) rotations about this point, scalings that leave the point fixed, and a reflection in a plane through this point.

The cases for U_{13}

The cases are described in detail in Krishnamurti and Earl (1992). There are two determinate cases which are illustrated in figure 8.

Case 1.1: There are two skew lines. Skew lines are not parallel; nor do they intersect at a point. The common perpendicular of two skew lines defines two distinguishable points,

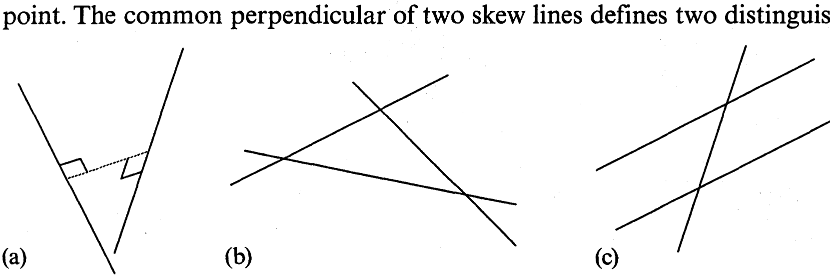


Figure 8. Examples illustrating the determinate cases for U_{13}: (a) case 1.1; and (b, c) case 1.2.

that is, the feet of this perpendicular on the lines. As two points are sufficient to determine a fixed scaling factor, an additional (distinguishable) point may be constructed on one of the lines. As such, this problem reduces to case 0.1, with three distinguishable points, for which a determinate number of possible valid transformations exist. A variant to this construction is described in Krishnamurti and Earl (1992).

Case 1.2: There are three coplanar lines not all parallel and not all concurrent at a single point. If no two lines are parallel, then the intersection points of these lines constitute three distinguishable points. Otherwise, the two points of intersection constitute two distinguishable points that can be augmented with a third point constructed as in case 1.1.

There are also three indeterminate cases as illustrated in figure 9.

Case 1.3: There are two nonparallel lines. From a base transformation, all other transformations result from scalings that leave the point of intersection fixed and a reflection in the plane defined by both lines.

Case 1.4: There are two (parallel) lines. The full set of transformations is generated by composing a base transformation with translations along the direction of the lines together with a reflection in a plane perpendicular to the parallel lines.

Case 1.5: There is a single line. The full set of possible transformations is found by composing the base transformation with rotations about the line (denoted as l), scalings that leave a point on l fixed, translations along the direction of l, and a reflection in a plane normal to l.

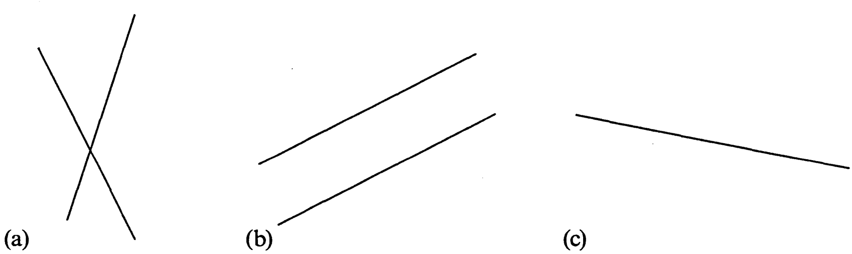


Figure 9. Examples illustrating the indeterminate cases for U_{13}: (a) case 1.3; (b) case 1.4; and (c) case 1.5.

The cases for U_{23}

The following cases are identified in Krishnamurti and Stouffs (1993). There is a single determinate case which is illustrated in figure 10.

Case 2.1: There are four planes not all parallel and not all lines of intersection are parallel, concurrent, or coincident. We can find two skew lines of intersection and, consequently, reduce the problem to case 1.1 for which there exists a determinate number of valid transformations.

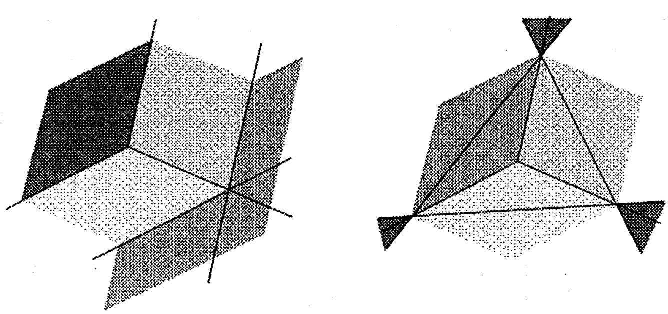


Figure 10. Examples illustrating the determinate case 2.1 for U_{23}.

There are five indeterminate cases in U_{23} which are illustrated in figure 11.

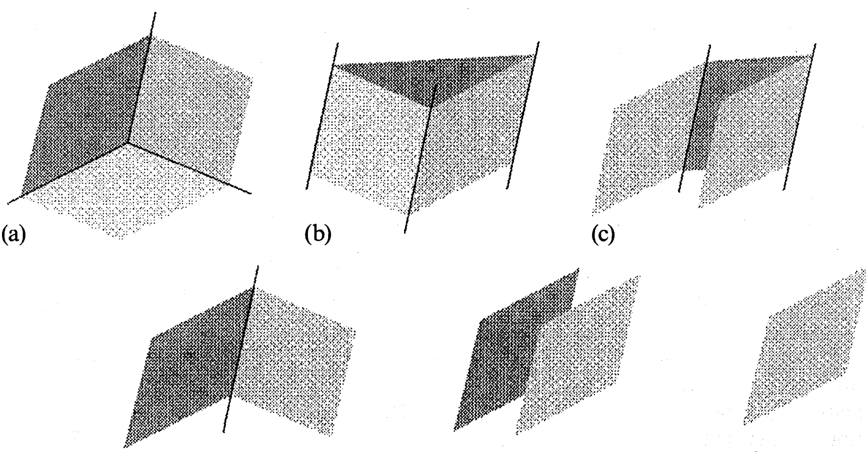
Case 2.2: There are three planes and not all the lines of intersection are parallel or coincident. Another way to formulate this is the following: *there are three planes, and their normal vectors are linearly independent.* All the lines of intersection are concurrent at a single point, and the problem reduces to case 1.3 for which there exists a possible indeterminate number of valid transformations (under scaling and reflection).

Case 2.3: There are three planes and these do not intersect in a single line. All the lines of intersection may be parallel but these do not coincide. This problem reduces to case 1.4 for which there exists a possible indeterminate number of valid transformations (under translation and reflection).

Case 2.4: There are two nonparallel planes. The full set of transformations is generated by composing a base transformation with translations along the direction of the line of intersection together with scalings that leave a point on this line fixed and a reflection in a perpendicular plane.

Case 2.5: There are two (parallel) planes. A base transformation maps the carriers of the two planes in a onto the respective carriers of two planes in s. All other transformations result from translations along two perpendicular axes parallel to these planes, rotations about a line normal to both planes, and a reflection in a plane through this line.

Case 2.6: There is a single plane. A base transformation maps the carrier of a plane in a onto the carrier of a plane in s. The full set of possible transformations is generated by composing this base transformation with translations along two perpendicular axes parallel to the plane, scalings that leave a point on the plane fixed, rotations about a line normal to the plane, and a reflection in a plane through this line.



(a) (b) (c) (d) (e) (f)

Figure 11. Examples illustrating the indeterminate cases for U_{23}: (a) case 2.2; (b, c) case 2.3; (d) case 2.4; (e) case 2.5; and (f) case 2.6.

Shape recognition revisited

In the preceding analysis all possible cases for each of the algebras U_{i3}, $0 \leqslant i \leqslant 3$, were considered. However, in many cases shape rule application is defined in a cartesian product of algebras. Earlier we defined shape rule application in an algebra V, $V = U \times V_0$, of labeled shapes. More generally, shape rule application can be defined in any cartesian product of algebras $U^1 \times U^2 \times \ldots$. As such, it is insufficient to enumerate the cases for each of the algebras U_{i3} separately. Below we consider the enumeration of the cases for any cartesian product of algebras U_{i3}, $0 \leqslant i \leqslant 3$.

Consider a shape s in the algebra $U_0 \times U_1 \times U_2 \times U_3$. The primary distinguishable elements can be augmented with constructed ones, such as the point of intersection of two lines, the line of intersection of two planes, the normal line to a plane through a point, and so on. Based on the combinations of independent transformations that yield the full set of possible valid transformations from a base transformation, only ten primary cases remain, of which one represents the determinate case and the remaining nine the indeterminate cases. We consider the *degrees of freedom* (DOF) corresponding to an indeterminate case to be the number of independent transformations (not including a possible reflection). Single transformations include a rotation about a line, a translation along a line, and a scaling. A single plane defines two independent translations; a general rotation about a single point constitutes three independent rotations. We use T to denote a translation, with T_l a translation along a line l, and T_u and T_v two (perpendicular) translations parallel to a plane. We use R to denote a rotation, with R_l a rotation about the line l, R_n a rotation about the normal n to a plane, and R_x, R_y and R_z rotations about the major axes. We use S to denote a scaling.

The ten determinate and indeterminate cases are summarized in table 4. Cases II through IV correspond to a single degree of freedom, either rotational, translational, or scaling. Cases V and VI have two DOF of which one corresponds necessarily to a scaling; the other is either rotational or translational. Cases VII and VIII have three DOF: a single distinguishable line results in a translation degree of freedom along the line, a rotational degree of freedom about the line, as well as a scaling; two parallel planes give way to a single rotational degree of freedom and two translational DOF. Cases IX and X have a maximal four DOF—either three rotational or one rotational and two translational, together with a scaling. No other combinations are possible: an axis of rotation corresponds to a single degree of freedom, and a centre of rotation to three DOF. No construct allows for only two rotational DOF. Similarly, any distinguishable element removes at least one translational degree of freedom; only a distinguishable plane allows for two translational DOF, but for only a single rotational degree of freedom.

Table 4. Ten possible combinations of independent transformations corresponding to one determinate (I) and nine indeterminate (II–X) cases for the subshape recognition problem in $U_0 \times U_1 \times U_2 \times U_3$.

Case	Distinguishable elements	Degrees of freedom	Rotation	Translation	Scaling
I	3 noncolinear points	0			
II	2 distinct points	1	R_l		
III	2 nonparallel lines	1			S
IV	2 parallel lines	1		T_l	
V	1 point and 1 line through the point	2	R_l		S
VI	1 line and 1 plane through the line	2		T_l	S
VII	1 line	3	R_l	T_l	S
VIII	2 parallel planes	3	R_n	T_u, T_v	
IX	1 point	4	R_x, R_y, R_z		S
X	1 plane	4	R_n	T_u, T_v	S

Each *primary* case defines a set of *secondary* cases, each of which can be reduced to the primary case by constructing additional distinguishable elements. However, it is computationally expensive to construct all possible distinguishable elements a priori, in order to determine the particular primary case. Therefore, we list below the secondary cases for each primary case, for all possible cases of combinations of primary distinguishable elements. Table 5 summarizes all nonredundant [6] cases with at least one distinguishable point. Table 6 summarizes all nonredundant cases without

Table 5. All nonredundant cases with at least one distinguishable point, based on combinations of primary distinguishable elements.

Distinguishable elements			Case	Degrees of freedom
points	lines	planes		
3, noncolinear			I.a	0
2		1, nonperpendicular	I.b	0
2			II.a	1
1	1, noncolinear		I.c	0
1	1, colinear	1, noncoplanar, nonperpendicular	I.d	0
1	1, colinear	1, coplanar	III.a	1
1	1, colinear		V.a	2
1		2, noncolinear, intersection	I.e	1
1		2, colinear, intersection	III.b	1
1		1, noncoplanar	II.b	1
1		1, coplanar	V.b	2
1			IX.a	4

Table 6. All nonredundant cases without distinguishable points but at least one distinguishable line, based on combinations of primary distinguishable elements.

Distinguishable elements		Case	Degrees of freedom
lines	planes		
3, coplanar, not all parallel nonconcurrent		I.f	0
2, skew		I.g	0
2, coplanar, nonparallel	1, nonconcurrent	I.h	0
2, coplanar, nonparallel		III.c	1
2, parallel	1, nonparallel	I.i	0
2, parallel		IV.a	1
1	2, nonconcurrent, not both parallel or perpendicular to the line	I.j	0
1	2, perpendicular to the line	II.c	1
1	1, nonparallel, nonperpendicular	III.d	1
1	1, parallel	IV.b	1
1	1, perpendicular	V.c	2
1	1, coplanar	VI.a	2
1		VII.a	3

[6] An example of redundancy would be two distinguishable points and one noncolinear line with at least one point: the second, possibly colinear, point is redundant as case I.c shows.

Table 7. All nonredundant cases with only distinguishable planes.

Distinguishable planes	Case	Degrees of freedom
4, not all parallel, not all intersections lines parallel or concurrent	I.k	0
3, not all intersection lines parallel	III.e	1
3, not concurrent in a single line	IV.c	1
2, nonparallel	VI.b	2
2, parallel	VIII.a	3
1	X.a	4

distinguishable points but at least one distinguishable line. Table 7 summarizes all cases with only distinguishable planes, which correspond to the cases for U_{23}. The cases listed below are numbered according to their primary case. Thus, case $d.n$ refers to the nth secondary case for the dth primary case.

Case I.a is identical to case 0.1; it applies to case 1.1, case 1.2, and case 2.1 as well. Case I.a represents primary case I.

Case I.a: There are three noncolinear points. We reduce each of the following cases to case I.a by constructing the necessary distinguishable points.

Case I.b: There are two points together with a plane not perpendicular to the line defined by both points. If at least one point is not coincident with the plane, then construct the foot of the perpendicular from this point onto the plane as a third distinguishable point. These three points are not colinear; if they were the plane would be perpendicular to the line through both points. If both points are coincident with the plane, then construct a point on the normal with the plane through one point, such that the distance to this point is identical to the distance between the two original points (figure 12).

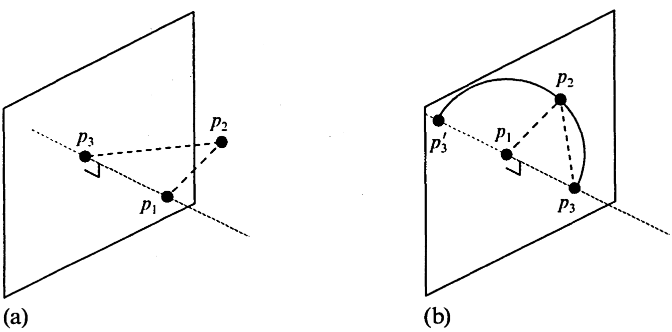


Figure 12. Illustrations of case I.b: (a) at least one point not conicident with the plane; and (b) both points coincident with the plane.

Case I.c: There is a point and a line not colinear with the point. Construct the foot of the perpendicular from the point onto the line and construct a third point on the line at equal distance from the foot as the distance between the foot and the first point, as shown in figure 13(a) (Krishnamurti and Earl, 1992).

Case I.d: There is a point, a line colinear with the point, and a plane neither coincident with the point nor perpendicular to the line. Construct the foot of the perpendicular from the point onto the plane. As this point is not colinear with the line, we have a situation similar to case 1.c. However, we can construct a third point on the line at an equal distance from the original point as the distance from the foot to this point [see figure 13(b)].

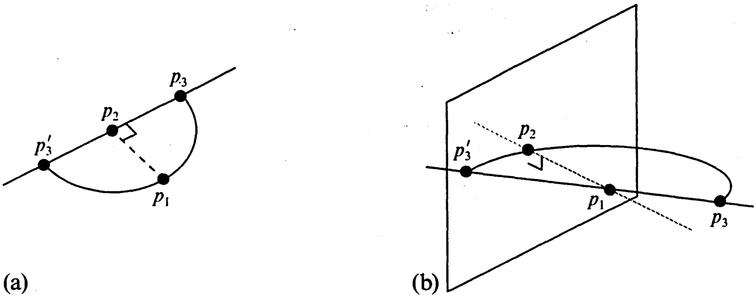


Figure 13. Illustrations of: (a) case I.c, and (b) case I.d.

Case I.e: There is a point together with two planes neither of which is coincident with the point. As the line of intersection is not colinear with the point, this problem is reduced to case I.c.

Case I.f: There are three coplanar lines not all parallel and not all concurrent at a single point. See case 1.2.

Case I.g: There are two skewed lines. See case 1.1.

Case I.h: There are two coplanar nonparallel lines together with a plane, which are not all concurrent at a single point. Consider the point of intersection of both lines together with one of the lines, which is not perpendicular to the plane, and the plane; this corresponds to case I.d.

Case I.i: There are two (parallel) lines and a plane which is neither parallel to the lines nor coincident with a line. Construct the intersection points of both lines with the plane. Construct a third point on one of the lines at equal distance from the point of intersection of this line with the plane as the distance between both intersection points [figure 14(a)].

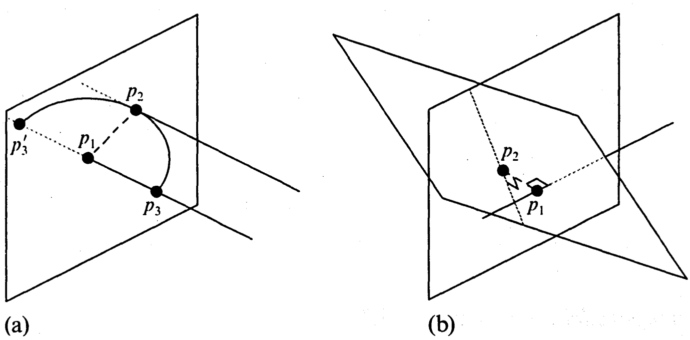


Figure 14. Illustrations of (a) case I.i, and (b) case I.j.

Case I.j: There is a single line together with two planes, not parallel or perpendicular to the line, such that these are not all concurrent at a single point. The construction is dependent on whether one of the planes is parallel to the line or not. If one plane is parallel to but not coincident with the line, and the other plane is neither parallel to nor coincident with the line, then the line of intersection of both planes is skew with respect to the original line [figure 14b)]. This reduces to case I.g or case 1.1. Otherwise, neither plane is parallel to nor coincident with the line, both planes are not perpendicular to the line, and all three elements are not concurrent at a single point. Consider the (two) points of intersection of the line with both planes together

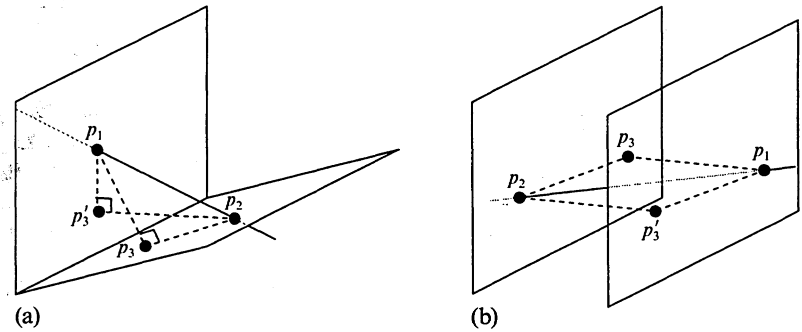


Figure 15. Illustrations of case I.j.

with the foot of the perpendicular from one of the intersection points onto the other plane. These three points cannot be colinear (figure 15).

Case I.k: There are four planes not all parallel and not all lines of intersection are parallel, concurrent or coincident. See case 2.1.

Case II.a specifies a single axis of rotation (without scaling); it is identical to case 0.2. It represents primary case II.
Case II.a: There are two (distinct) points. We reduce each of the following cases to case II.a by constructing the necessary distinguishable points.

Case II.b: There is one point, and one plane not coplanar with the point. Construct the foot of the perpendicular from the point onto the plane. The resulting axis of rotation (defined by both points) is perpendicular to the plane, such that the plane is mapped onto itself under the transformation.

Case II.c: There is one line, and two planes perpendicular to the line. The two points of intersection of the line with the planes are distinct.

Cases III.a through III.e specify a scaling. Case III.c is the representative case; it is identical to case 1.3 and applies also to case 2.2. We reduce each of the following cases to case III.c by constructing the necessary distinguishable lines.
Case III.a: There is one point, one line, and one plane, all coincident. Construct a second line perpendicular to the first line, coincident with both the point and the plane.

Case III.b: There is a point, and two planes coincident with the point. Construct the foot of the perpendicular from the point onto the plane. The resulting axis of rotation (defined by both points) is perpendicular to the plane, such that the plane is mapped onto itself under the transformation.

Case III.c: There are two nonparallel lines. See case 1.3.

Case III.d: There is one line and one plane; these are neither parallel nor perpendicular. Construct the line of intersection of a second plane, coincident with the line and perpendicular to the first plane, with the first plane.

Case III.e: There are three planes and not all the lines of intersection are parallel or coincident. See case 2.2.

Case IV.a specifies a single direction of translation (without scaling); it is identical to case 1.4 and applies also to case 2.3. Case IV.a represents primary case IV.
Case IV.a: There are two parallel lines. We reduce each of the following cases to case IV.a by constructing the necessary distinguishable lines.

Case IV.b: There is one line and one parallel plane. There exists a unique second line coincident with the plane and parallel to the first line, such that the foot of the perpendicular from any point on the first line lies on the second line (that is, the second line constitutes a normal projection of the first line onto the plane).

Case IV.c: There are three planes and these do not intersect in a single line. See case 2.3.

Case V.a, which represents primary case V, specifies a single axis of rotation with scaling.
Case V.a: There is a single point, and a line coincident with the point. The line constitutes the axis of rotation, and the single point inhibits any translation but allows for a scaling (Krishnamurti and Earl, 1992). We reduce each of the following cases to case V.a by constructing the necessary distinguishable line and/or point.

Case V.b: There is a single point, and a plane coincident with the point. The line normal to the plane and coincident with the point constitutes the axis of rotation.

Case V.c: There is a single line, and a plane perpendicular to the line. The line constitutes the axis of rotation, and the point of intersection of the line and the plane constitutes the fixed point.

Case VI.a, which represents primary case VI, specifies a single direction of translation with scaling.
Case VI.a: There is a single line, and a plane coincident with the line. The line defines the direction of translation, and the plane inhibits any rotation about the line but allows for scaling.

Case VI.b: There are two nonparallel planes. See case 2.4.

There is only one case, case VII.a, which specifies a single axis of rotation, a single direction of translation, and a scaling. It is identical to case 1.5.
Case VII.a: There is a single line.

There is only one representative case, case VIII.a, which specifies a single axis of rotation and two (perpendicular) directions of translation (without scaling); it is identical to case 2.5.
Case VIII.a: There are two (parallel) planes.

Case IX.a (the representative of primary case IX) specifies three (perpendicular) axes of rotation and a scaling; it is identical to case 0.3.
Case IX.a: There is a single point.

Case X.a (the representative of primary case X) specifies a single axis of rotation, two (perpendicular) directions of translation, and a scaling; it is identical to case 2.6.
Case X.a: There is a single plane.

This completes the enumeration of the secondary cases.

Variations on a common theme

The classification, according to the number and type of DOF, is based solely on distinguishable elements. That is, only the carriers of the shape segments and not the segments themselves define the specific classification. Each class is characterized by a base transformation, from which all possible transformations can be derived in correspondence to the DOF of this class. A possible transformation is a transformation that maps the given shapes under the part relation. As such, the part relation defines a constraint on the possible transformations.

Gero and Yan (1993; 1994) consider a notion of emergent shapes that are visually recognized as such, which they refer to as *emergent visual shapes*. In particular, using shapes in U_1, they consider as an emergent visual shape any shape that is embedded in the carriers of the given shape and the segments of which are bounded by the points of intersection of the carriers.[7] These carriers and their points of intersection define precisely the distinguishable elements used to determine the class of transformations under shape recognition. The resulting set of possible transformations is constrained by the embedding of the segments of the transformed shape in the carriers (instead of the segments) of the given shape.

Emergent shapes may be constrained by, for example, topological, geometrical, or dimensional considerations. For instance, Gero and Yan require emergent visual shapes to be simple closed polygons. A simple variant of the subshape recognition algorithm (Krishnamurti, 1981) can detect general emergent visual shapes. Stiny (1977; 1980) characterizes shape recognition as dependent on a set of parameters specified on the emergent shape, together with the constraints or bounds on the parameters. Other requirements include emergent shapes augmented with attributes such as labels or weights. The enumeration of the recognition cases for labeled (and weighted) shapes can proceed along the lines described in this paper.

Acknowledgements. We would like to thank George Stiny for his helpful comments.

References

Bridges A H, 1991, "DAC or design and computers", in *CAAD Futures '91* Ed. G Schmitt (Vieweg, Wiesbaden) pp 65–76
Carr E H, 1961 *What is History?* (Vintage, New York)
Chase S C, 1989, "Shapes and shape grammars: from mathematical model to computer implementation" *Environment and Planning B: Planning and Design* **16** 215–242
Galluzzi P (Ed.), 1987 *Leonardo Da Vinci, Engineer and Architect* (Montreal Museum of Fine Arts, Montreal)
Gero J S, Yan M, 1993, "Discovering emergent shapes using a data-driven symbolic model", in *CAAD Futures '93* Eds U Flemming, S Van Wyk (Elsevier, Amsterdam) pp 1–17
Gero J S, Yan M, 1994, "Shape emergence by symbolic reasoning" *Environment and Planning B: Planning and Design* **21** 191–212
Hocking J G, Young G S, 1988 *Topology* (Dover, New York)
Krishnamurti R, 1981, "The construction of shapes" *Environment and Planning B: Planning and Design* **8** 5–40
Krishnamurti R, Earl C F, 1992, "Shape recognition in three dimensions" *Environment and Planning B: Planning and Design* **19** 585–603
Krishnamurti R, Stouffs R, 1993, "Shape grammars: motivation, comparison and new results", in *CAAD Futures '93* Eds U Flemming, S Van Wyk (Elsevier, Amsterdam) pp 57–74
Leonard H S, Goodman N, 1940, "The calculus of individuals and its uses" *The Journal of Symbolic Logic* **5** 45–55
Mitchell W M, 1993, "A computational view of design creativity", in *Modeling Creativity and Knowledge-Based Creative Design* Eds J S Gero, M L Maher (Lawrence Erlbaum, Hillsdale, NJ) pp 25–42
Oxman R E, Oxman R M, 1991, "Refinement and adaptation: two paradigms in form generation in CAAD", in *CAAD Futures '91* Ed. G Schmitt (Vieweg, Wiesbaden) pp 313–328
Stiny G, 1977, "Ice-ray: a note on Chinese lattice designs" *Environment and Planning B: Planning and Design* **4** 89–98
Stiny G, 1980, "Introduction to shape and shape grammars" *Environment and Planning B: Planning and Design* **7** 343–351
Stiny G, 1982, "Shapes are individuals" *Environment and Planning B: Planning and Design* **9** 359–367
Stiny G, 1991, "The algebra of designs" *Research in Engineering Design* **2** 171–181

[7] Gero and Yan refer to the carriers of a given shape as infinite maximal lines. The concept of a 'carrier' expresses independence of dimensionality and generalizes emergent visual shapes to shapes that are bounded by lines (or planes, volumes, etc) of intersections of the carriers.

Stiny G, 1992, "Weights" *Environment and Planning B: Planning and Design* **19** 413–430
Stiny G, 1993a, "Emergence and continuity in shape grammars", in *CAAD Futures '93* Eds U Flemming, S Van Wyk (Elsevier, Amsterdam) pp 37–54
Stiny G, 1993b, "Boolean algebra for shapes and individuals" *Environment and Planning B: Planning and Design* **20** 359–362
Stiny G, 1994, "Shape rules: closure, continuity and emergence" *Environment and Planning B: Planning and Design* **21** 549–578
Stiny G, 1996, "Useless rules" *Environment and Planning B: Planning and Design* **23** 235–237
Stouffs R, 1994 *The Algebra of Shapes* PhD thesis, Department of Architecture, Carnegie Mellon University, Pittsburgh, PA