CLOSING IN ON AN OPEN PROBLEM

reasons and a strategy to encode emergent subshapes

Milton Tan, RIBA
Graduate School of Design
Harvard
&
School of Architecture
National University of Singapore

ABSTRACT

The interpretation of drawings, by breaking them into subshapes and classifying these subshapes, is an essential part of creative designing. Drawings must be open to different interpretations – i.e. different decompositions into parts, and classification of these parts in different ways – but conventional CAD systems do not readily allow this. Their data structures are too inflexible, and they do not provide subshape or implied shape recognition capabilities.

This paper discusses the centrality of emergent forms in the design process and proposes a data-structure based on construction lines and ordered lists which enables shapes as collection of lines and arcs to be efficiently encoded. The strategy to build a design tool around this data structure is also presented.

DESIGN – AN OPEN PROBLEM

One can acquire certainty only by amputating inquiry.

Creative thinking involves originality, innovation and new insights which frequently contradict and challenge existing beliefs and practices; a sharp contrast to problem-solving as a goal-directed activity within well-defined 'problem spaces', solvable by "a knowledgeable man without the need for further information" (Rittel, 1972, p.392), and not expected to yield surprising solutions. Creativity presupposes that problem-definitions are neither beyond contention nor exclusive; rather, they must be open to new interpretations and extendable2. Moreover, these new considerations can affect and even invalidate prior knowledge – questioning monotonicity assumptions. In other words, problems can modify its own course or be modified.

It is beyond the scope of this paper to fully review the issue of open problems but an example should serve well to highlight its key processes in order to focus on its specific implications for

1 See Newell and Simon, 1972.

2 Even in the sciences, no 'fact' can remain beyond controversy (e.g., see Punnam, 1987). If it were so, no new theories would be possible; in fact no theory could ever have been formed.
architectural design. Max Wertheimer (1959, pp.266-268) discusses a "stained-glass window puzzle"—to determine the area of the shaded part in Figure 1, given the diameter of its circle:

![Figure 1: Given the diameter of the circle, what is the area of the shaded part?](image)

In the problem-solving paradigm, a typical response to this puzzle would be to identify the geometry of the shaded area and its corresponding area-calculation formula, and subtract its area from the area of the (unshaded) circle. But a more creative approach would be to reconceptualize the figure as a circle within a shaded square with semi-circles on two opposite sides (Figure 2a), moving the two semi-circles into the center (Figure 2b), and consider the original shaded part as a square (Figure 2c) whose area is simply the square of the given diameter.

![Figure 2: Solution by reconceptualization and transformation.](image)

The 3 stages illustrated in Figure 2 conforms with Kenneth Craik's model of thought (1943). He proposes that in thinking, 1) a process of "translation" encodes the 'external world', 2) a process of "transformation" operates on the encoded symbols, and 3) a process of "retranslation" returns the result of thinking to the 'external world' (Figure 3). In design, the corresponding processes are often known as conceptualization, transformation and representation. This paper is concerned with the first of these.

![Figure 3: Craik's model of thought (after Wade, 1977).](image)

Whilst a problem can always be reframed or reconceptualized in some other way, it is no guarantee of a better solution. In many practical situations, the basis of problems are seldom questioned.
because optimal solutions — or *maximizations* — are not crucial. However, reconceptualization is often the only means to progress and to escape from 'programmed' stereotypes on the one hand and randomly-generated solutions on the other.

It might seem anarchic to suggest that a destination should depend on the route chosen, a solution to depend on how the problem is posed, or an answer on the question asked; but in design exploration, it is desirable — even vital — to recognize that its intermediate representations can be *open to new and different interpretations* and therefore leading to new lines of inquiry. For instance, how would one *describe* the shape in Figure 4?

![Figure 4: A simple cross shape?](image)

The Gestaltist would naturally say that it is simply two intersecting rectangles; according to the Gestalt law of "good continuation" we perceive the composition that preserves the fewest lines. But the description, surely, depends on how one chooses to carve up the shape, however unpopular, unlikely or 'illegal' that may be; e.g., in Figure 5, the more interesting decompositions are unlikely to found by consensus.

![Figure 5: A few reconceptions of Figure 4.](image)

---

3 Herbert Simon coined the term "satisficing" for this sub-optimal behavior which contradicts the notion of maximization in expected Utility Theory. Perkins (1981, p.60) put it in human terms — "Search...typically satisfies rather than maximizes because typically one can't maximize, doesn't know how, or wouldn't gain overall".

4 For example, see Wertheimer (1923), Kohler (1929) and Anheim (1974).
To 'see as' is, simply, to conceptualize — to ascribe different things as the same by equivalence; to group objects (or events) into classes and respond to them in terms of their class membership rather than uniquely5. It is ultimately about making a distinction of similarities to defy eventual differences; to check perceptual entropy as it were. In fact, conceptualization is not an option, for it makes no sense to describe human perception free from conceptual contamination. Bruner, Goodnow and Austin exposes the irony of a fully utilized discriminating capability as "infinitely enslaved to the particular" (Bruner et al., 1956).

It is not sufficient to have the whole world at one's disposal — the very infinitude of possibilities cancels out possibilities, as it were, until limitations are discovered (Sessions, 1956; p.31).

Therefore, what one finds depends on what one is looking for; "extensive search accomplishes nothing unless the person looks in the right places" (Perkins, 1981, p.132). Putting it another way, the correctness of an solution is not independent of how the problem is framed, or reframed. In design, given a shape conceptualized in a particular way, transforming and representing it is a relatively straightforward (problem-solving) task — the constituent parts and their essential properties being defined by the concept. But concepts, especially in creative activity, are seldom either 'given', complete or beyond contention. The creative mind (or any mind for that matter) does not inhabit a ready-made world and is therefore free to seek or modify problems, and not just to solve them. Choosing to see the shape in Figure 4 in the few ways suggested in Figure 5 require the use of different conceptual filters. These ways of seeing — like Nelson Goodman's "world-making" (Goodman, 1978) or Mitchell's "universes of discourse" (1990a) — determine the developmental potential of the shape and therefore its emergent forms. In terms of rational thought, this contamination is often considered undesirable or transitional at best; but for creative thought, it is indispensable. In architectural design, the choice of conceptual filters applied to an initial shape — a parti — (such as Figure 4) will result in dramatically different designs even though only a determinate set of transformation and representation rules are available.

Before taking a closer look at the function of conceptual filters, the phenomena of nonmonotonicity in design needs to be briefly mentioned. Rational thinking, along the lines of traditional first-order logic, is committed to the monotonicity assumption: "the addition of new facts and rules to the critical knowledge base does not falsify any fact already in the knowledge base" (Mitchell, 1990a, p.80). But the notion of open problems which demands that no fact is made 'sacred' or unequivocally related to other facts past, present or future, presupposes that monotonic logic cannot always be applied in creative design thinking. The result of a transformation on a part of a design in progress conceptualized then in a particular way is returned to the rest of the work for further conceptualization; its new presence can of course change the status quo; indeed, for it to be a creative move, it must. For a simple example, consider the parametric variation of the openings in a wall as shown in Figure 6a:

![Figure 6: A wall with 2 parametrically variable openings.](image)


6 The converse is of course not necessarily true.
By simply rescaling the size of the openings, prior assumptions about the wall is affected. In Figure 6b, the ‘wall’ between the openings is better treated as a column or pier; in Figure 6c, it reduces to a post or frame, the ‘wall’ above the openings may be considered a beam.

CONCEPTUALIZATION CAPACITIES

Before dealing with the issue of conceptual filters in designing, it is necessary at this juncture to briefly review three important conceptualization capacities in order to isolate types of emergent forms which would interest designers. They are abstraction, completion, and interpretation.

1 Abstraction

Abstraction accounts for the fact that a full description of an object (or event) can consist of a set of relatively independent (albeit associated) descriptions, one for each part or property of the object (Rock, 1983, pp.68-69). More than that, we abstract with our finite perceptual capacities only those things which are noteworthy. We have already seen in Figure 5 how very different subshapes can be abstracted form the same figure. To be pedantic about it, abstractions (such as subshapes) must be sub-sets of the object. For a drawing taken as a collection of lines, this means that all combinations of line segments within maximal lines qualify as subshapes (Stiny, 1980); technically, they can be any fragment of the maximal line and need not be delimited by intersections, although, perceptually, these are more apparent and generally more meaningful (Figure 7).

![Figure 7: Subshape qualification at intersections (a) and between intersections (b), (c) and (d).](image)

2 Completion

Completion (or, more informally, making-up) is the answer — at least in part — to our sense limitations. It is amazing how little information and time we actually need (or have, depending how you look at it) to form reasonably accurate deductions. In Figure 8, we readily make up for the letter “A” for “CAT” because “CHT” would not have made much sense; but we hardly think of touching the letter “H” in “THE” because “TAE” would be equally senseless.

![Figure 8: Making up for what would make more sense.](image)

Another type of completion can be seen in ‘low-pass’ images — we have the proclivity to interpolate the data into a finer-grained, and therefore more complete representation. Although it is a rather extreme case of image completion, it must be by a similar process that we ordinarily recognize faces and read handwriting with relative ease.
The clues can be even more subtle – they may (intentionally or otherwise) suggest forms which require a higher degree of abstraction – of selecting some key features of the stimuli and making up others in order to qualify the fabrication as a member of a different class of things. The first example of this is Kanizsa’s (1955) ‘phantom’ figure, Figure 9, which induces a secondary emergent shape.

![Figure 9: Kanizsa ‘phantom’ figure.](image)

3 Interpretation

Abstraction and completion set the stage for further perceptual complexities because an abstraction or completion is rarely made autonomous but is instead often returned to its host context (or another context) after some intentional transformation.

Consistency is the main concern with transformations, i.e. the effects of the transformation to a context. Take for example Jastrow’s duck-rabbit figure in Figure 9 – depending on whether the projecting appendage is labelled “ears” or “beak”, the naming of the remaining figure (including its orientation) rallies around this hypothesis to attempt a consistent interpretation of either a rabbit (facing right) or a duck (facing left). This type of ambiguity is a fact of life; not a figment of the imagination. Ludwig Wittgenstein, in Philosophical Investigations (Wittgenstein, 1968), was impressed by the fact that the duck-rabbit figure can have two valid (but mutually exclusive) interpretations. It is a well-established fact that the human imagination is capable of finding recondite or unlikely relationships between seemingly disconnected things, then mustering support within itself to self-justify them, sometimes against all odds or conventions!

![Figure 10: The Duck-Rabbit ambiguous figure.](image)

There have been many studies of ambiguous – or more generally, multistable or equivocal forms – for example, Stephen Kosslyn in Image and Mind (1980), Ronald Finke in Principles of Mental Imagery (1989), Rudolf Arnheim in Art and Visual Perception (1974), Irvin Rock in The Logic of Perception (1987), and Mark Fineman in The Inquisitive Eye (1981). They all generally point to the fact that although there are some preferred ways, shapes can often be conceptualized in many other ways.

Gestalt psychologists have long maintained that there are ‘natural’ and more immediate recognition of certain types of shapes over others. Even if this is true, there is no compelling reason why these
should be preferred by designers; in fact a creative tendency would be to seek 'non-gestalt' emergent shapes! and leaving little room for 'customary practice' or 'consensus judgement'.

DESIGNING WITH CONCEPTUAL FILTERS

Conventional CAD systems, based on a 'tinker-toy' approach of providing a priori primitives for accumulating shapes to represent designs, clearly do not cope with the sorts of conceptual capacities discussed above which are concerned with partial transformation of designs in progress. Specifically, they do not provide the following:

1. Enable the abstraction of subshapes.
2. Support the completion of implied (or 'phantom') shapes.
3. Allow multiple interpretation (or recognition) of ambiguous shapes.

Creativity has to do partly with the conceptualization of emergent forms — the ability to filter out significant aspects of a design in progress; aspects which are not limited to the overt interventions by the designer but include 'side-effects' described above. Traditionally, this is manifested in the use of the ubiquitous tracing paper — a designer would isolate an emerging form to a new working 'plane' by tracing it off its wider and often ambiguous or multi-stable context. After it has been worked on, its transformed state may be returned to its original context for overall evaluation and further conceptualization. In view of this, the tracing paper is the premier design tool, less the often cited T-square or the pencil which are primarily tools of representation.

A design tool which copes with these three demands would require, inter alia, a fundamentally different approach in handling its form of representation.

WHAT'S IN A SHAPE?

A shape is taken as a representation of the boundaries or edges of an object. As with any complex problem, it is prudent to partition the representation of shapes into more manageable portions. For the present study, it is taken as collections of straight lines in a two-dimensional plane. Little would be gained at this point of the inquiry by redrawing the battle lines to include three-dimensional shapes with complex geometries; they are in a continuum and do not invalidate each other.

Emergent forms require the abandonment of the assumption that shapes consist of indivisible or 'atomic' parts. But this leads immediately to severe combinatorial explosion — a line can contain an infinite number of sub-lines. But then again, designers do not usually perform exhaustive search. Instead, they conceptualize. They know ahead of time the sort of things which interests

---

7 This representation can of course be very different in different 'worlds', e.g., Aronheim (1974) distinguishes between physical, perceptual and visual shapes.
8 Shape Grammar research has basically taken this approach. See Siny, 1980a; 1990.
9 There is of course no problem if a line is never to be sub-divided in its lifetime, but, as we have seen, emergent shapes cannot be made out of indivisible lines except in highly contrived cases where the shape is made out of the maximum number of line segments.
them; these biases, attention mechanism or conceptual filters enables comprehension of what otherwise are effectively meaningless bits of sensory data.

SHAPE REPRESENTATION

The basic conditions which produces emergent subshapes are line 'colinearity' and intersection. In the first case, when two lines are colinear, three types of 'emergent' line types may be formed: maximal lines, sub-lines and phantom lines.

1 Maximal Lines

If any two unequal colinear lines, (a b) and (c d), share an end point or if at least one end point of a line is coincident with the other line, the two lines can be reduced to a single maximal line (Figure 11):

```
1  a  b
  c  d

2  a  b
  c  d

3  a  b
  c  d

4  a  b
  c  d
```

Figure 11: Pairs of colinear lines (slightly separated for legibility) which can be reduced to a maximal line.

The maximal lines from Figure 11 can be represented by their pair of end-points:

- Line 1: (a b)
- Line 2: (a b)
- Line 3: (a d)
- Line 4: (a d)

But the intermediate points should also be included as they are potentially significant to the shape and would need to be recalculated every time they are needed. Without resorting to long lists of separate colinear lines (but admittedly with some compromise to elegance) the maximal lines can be expressed more usefully as the set union of the lists (a b) and (c d):

- Line 1: (a d b)
- Line 2: (a c d b)
- Line 3: (a c b d)
- Line 4: (a b d)

Notice that the first and the last elements of the maximal line list denote its endpoints. To determine if two lines are connected, simply check if one of end point is common to both; e.g. applying a "set-exclusive" function on the first and last elements of the two lists should result only
in two elements. Similarly, intersecting lines share an intermediate point; conversely, when two lines intersect, their point of intersection is appended to both lists as intermediate points.

2 Sub-Lines

Further, the combination of pairs of all points in a maximal line list produces its discrete sub-lines (Figure 12):

```
 a b c d
```

```
 a b
 b c
 c d
```

```
 a c
 c d
```

```
 b d
```

Figure 12: Discrete sub-lines of maximal line (a b c d).

These can be considered 'first order' emergent lines. From combinations of 'first order' emergent lines, all subshapes (e.g. Figure 5) defined by end or intermediate points can be exhaustively enumerated.

3 Phantom Lines

But, in order to account for implied or 'phantom' shapes (such as Figure 9), at least three other orders of emergent lines have to be considered. Firstly, 'second order' emergent lines are produced from the pair-combinations between all points of two colinear maximal lines, Figure 13:

```
 a b c d e
```

```
 a d e
```

```
 b d e
```

```
 c d e
```

Figure 13: Second order emergent lines from colinear lines (a b c) and (d e).

Next, 'third order' emergent lines are produced from the pair-combinations between all points of a maximal line to other 'colinear' points (Figure 14).
These 'colinear' points can be expressed conveniently as points coincident with a construction line which is colinear with the maximal line. The intersection of a construction line with a maximal line will therefore produce an intermediate point on the maximal line; and the intersection of two construction lines will produce a 'free' point (Figure 15a). Figure 15b shows some simple examples of second and third order emergent shapes:

To complete the range of emergent shapes based on construction lines, 'fourth order' emergent lines are simply lines connecting 'free' points along construction lines. As these are not directly connected to any line segments, their presence is usually least apparent, but nevertheless can be significant as shown in Figure 16.
Figure 16: Fourth order emergent shapes – the two squares in b – formed at the intersections of construction lines.

Because of the significance of second and third order emergent shapes, conventional encoding of lines as a pair of endpoint coordinates is indifferent – if not unsuitable – to their efficient recognition and transformation. The problem is considerably compounded when other primitives, such as arcs, are introduced to the universe of possible shapes. Moreover, in terms of using conceptual filters to recognize shapes, parts of shapes are often naturally built up of other shapes, in addition to primitives such as lines and arcs; it is unreasonably reductionist and, ultimately, hopelessly inefficient to treat concepts of shapes purely as collections of primitives. As a first step away from this grid-lock, it is advantageous that points be used to define the relationships of lines, arcs and other primitives used to describe shapes, rather than passively fixing independent coordinates (Tan, 1990). This approach respects the fact that tokens of design are not autonomous and that their relationships are important to designers.

\textit{ecart} \footnote{"ecart" is French for "separation" or "deviation"; it also happens to be "trace" spelt backwards! The name captures the major theme of the program – design development by separating emerging forms of interest using conceptual filters.}

\textit{ecart} is a prototype CAD program based on the hypothesis that an essential part of creative designing involves an ongoing interpretation of drawings which are open to different conceptualizations – i.e., drawings which can be decomposed and classified into different combination of parts. It is an attempt to parallel the designer’s perception of emergent forms.

The shape recognition strategy includes:

1. A data structure, based on the idea of construction lines, that supports efficient recognition of emergent shapes by recording alignment relationships.

Currently, a shape is taken to be a collection of straight lines, arcs and other shapes. Lines and arcs are ordered lists of coincident points; points are lists of intersecting construction lines or circles; construction lines and circles are lists of their geometry-defining parameters:
Sets of 'conceptual filters' which direct attention to the types of emergent shapes that are relevant to particular contexts (and avoid combinatorial explosion); these can range from relatively simple filters such as "squares" or "ionic columns" to more complex ones such as "to see as a structural engineer", "to see as a space planner" or "to see as Palladio".

A shape type is defined by specifying its essential parts and their relationships. A particular combination of the parts which satisfies the relationship predicates qualifies as an instance of that type.

Defining conceptual filters as hierarchies of shape types and subtypes, thereby maintaining graded degrees of shape resemblance. Compound shapes are progressively 'unpacked' for subordinate specifications to recognize more 'unlikely' emergent shapes; e.g. four connected lines can be recognized as a "quadrilateral"; a "quadrilateral" with two pairs of parallel lines as a "rectangle"; a "rectangle" with equal sides as a "square".

This is a search strategy which attempts to model the ease of recognizing more specific instances but treating them as accidental instances of a less specific type by abandoning their essential properties; i.e. a "square" is an instance of a "rectangle", which is an instance of a "quadrilateral", which is an instance of four lines.

The current version of ecart provides an interactive interface of delineating construction lines and circles by two points; line segments or arcs are defined by picking on end points. The system automatically maintains the lists of maximal and intersecting lines and arcs (as described above). A search for an emergent shape is conducted by analyzing the parts, variables and relations specifications of the shape type as defined in an interactive 'modeless dialog'. Figure 18 shows a composition after Gottschaldt (1926) using ecart; the thin lines are construction lines.

Figure 19 shows a hidden 6-sided figure – different from the one intended by Gottschaldt11 – recognized in accordance with the terms of its conceptual filter where one was never explicitly specified. Note that, like 'phantom' shapes, one side had to be made up to complete this figure.

11 A hexagon.
Figure 18: A screen-dump from esart showing a composition after Gottschaldt.

Figure 19: A screen-dump from esart showing the use of a conceptual filter to recognize a hidden shape.
CONCLUSION

David Hubel once said that we still hardly know what visual signals are, let alone how those signals make visual sense. This project, beginning with ecart, is in a sense saying a similar thing of designing – we scarcely understand the nature of shapes in drawings. Designing is more than the solving of well-defined or closed problems where 'what you see is what you get'; on the contrary, design is often an open problem where 'what you see as is what you get'. ecart is an attempt to define a basic tool – conceptual filters – which parallel what designers do with tracing paper.

REFERENCES


