Perception of structure 2: Impossible Objects

(Aaron Sloman, University of Birmingham)

This is a sequel to the presentation of a challenge to vision researchers here:

http://www.cs.bham.ac.uk/research/projects/cogaff/challenge.pdf (Feb 2005)

That presentation complains that current AI work on vision is very limited in scope and omits investigation of such things as:

- **Perception of structure (at different levels),** e.g.
  - perception of 3-D parts and their relationships
  - Perception of motion in which relationships between parts of one object and parts of other objects change, including things like sliding along, fitting together, pushing, twisting, bending, straightening, inserting, removing, rearranging.
- **Perception of positive and negative affordances and causal relations,** e.g.
  - Possibilities for action, for achieving specific effects
  - Obstructions to action, and limitations of actions

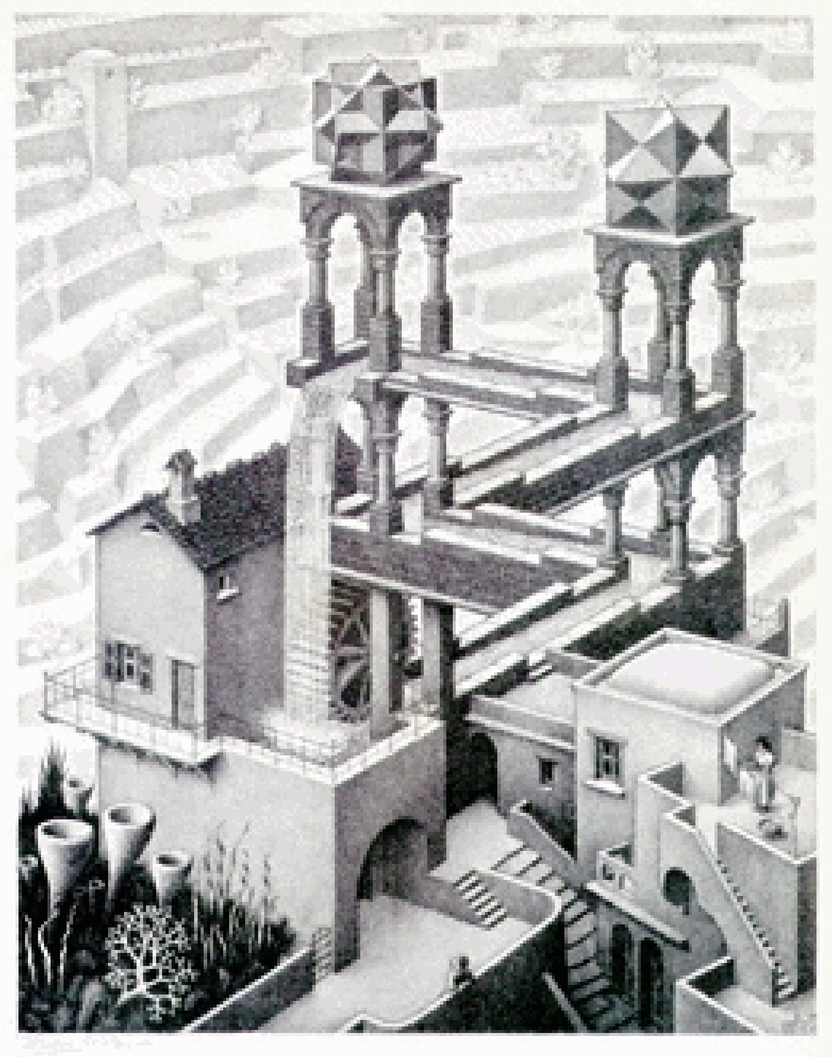
    especially as regards parts of complex objects, which can be grasped, pulled, pushed, twisted, rotated, squeezed, stroked, prodded, thrown, caught, chewed, sucked, put on (as clothing or covering), removed, assembled, disassembled, and many more...;

    and many variations and combinations of each of the above.

I asked some leading vision researchers to comment, but nobody was able to refute the main negative claims. In October 2005 I produced a set of requirements for human-like visual systems as being essentially concerned with perception of PROCESSES – also far from current AI/Robotic vision systems: http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0505

This sequel discusses some detailed requirements for visual mechanisms related to how typical (adult) humans see pictures of ‘impossible objects’.

Escher’s Weird World



Many people have seen this picture by M.C. Escher:

a work of art, a mathematical exercise and a probe into the human visual system.

You probably see a variety of 3-D structures of various shapes and sizes, some in the distance and some nearby, some familiar, like flights of steps and a water wheel, others strange, e.g. some things in the ‘garden’.

There are many parts you can imagine grasping, climbing over, leaning against, walking along, picking up, pushing over, etc.: you see both structure and affordances in the scene.

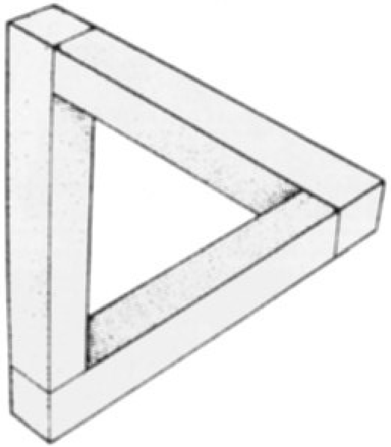
Yet all those internally consistent and intelligible details add up to a multiply contradictory global whole. What we see could not possibly exist.

There are several ‘Penrose triangles’ for instance, and impossibly circulating water.

Can you see the contradictions? They are not immediately obvious.

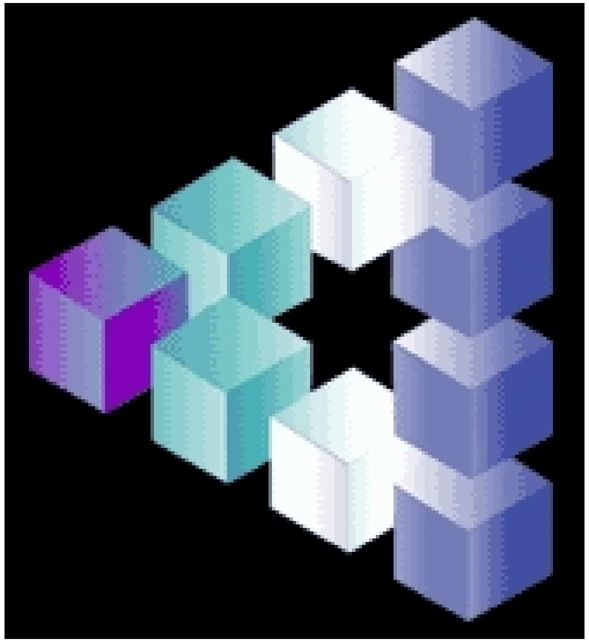
Perceiving impossible things

What is perception of spatial structure?



Pictures of ‘impossible objects’ tell us things about the nature of vision. The ‘Penrose Triangle’ is one of many depictions of impossible 3-D objects, often presented as relevant to the study of human vision, though implications spelled out below are rarely stated.

The Swedish artist Oscar Reutersvärd discovered impossible triangles earlier, in 1934, in the lower form on the right.



Notice how each little cube looks normal: you can easily visualise grasping any of them from different directions, pulling or pushing them out of the 'triangle' in various ways, etc. Yet the whole 3-D configuration is geometrically impossible. Why?

Such pictures refute the claim in Wittgenstein’s *Tractatus Logico-Philosophicus* (1921):

> 3.0321 Though a state of affairs that would contravene the laws of physics can be represented by us spatially, one that would contravene the laws of geometry cannot.

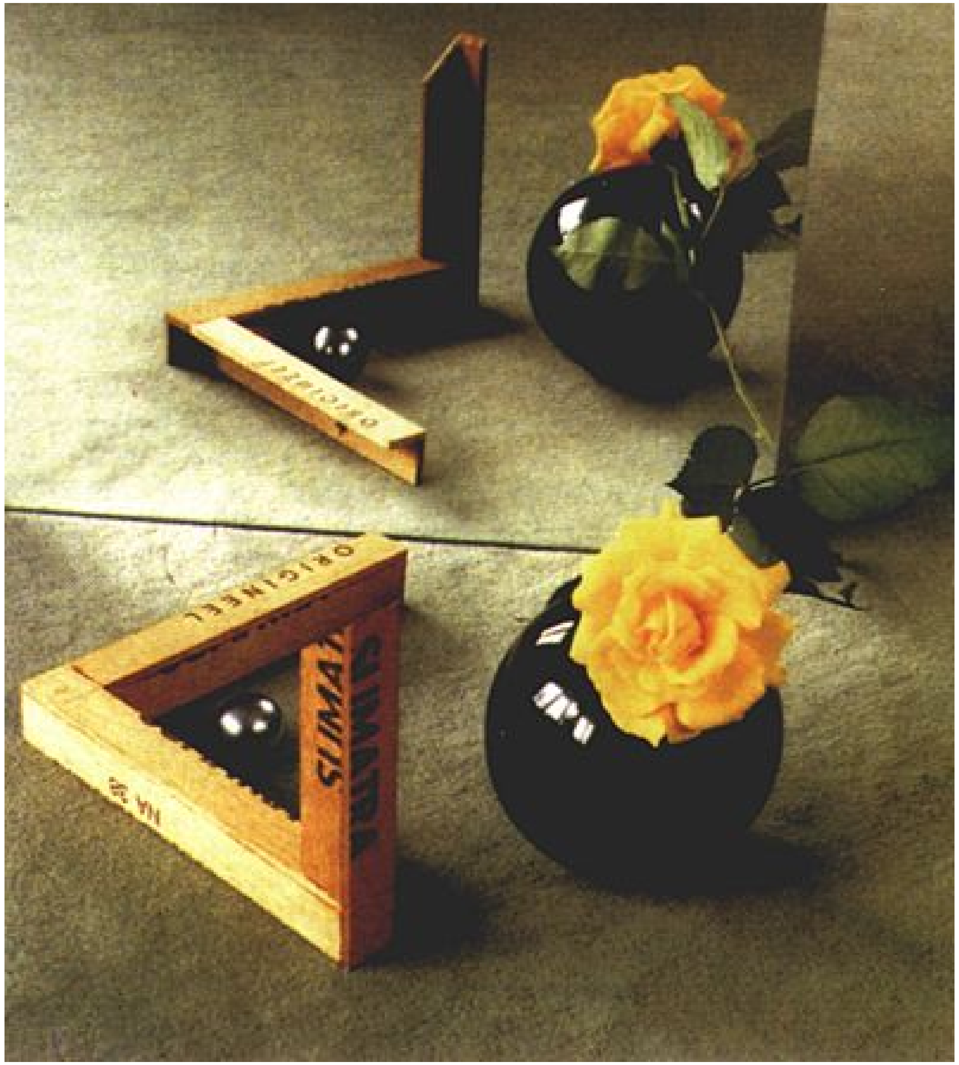
I once upset a philosophy research seminar discussing that assertion when I was a DPhil student in Oxford around 1960, by demonstrating that it is very easy to draw a picture of a round square, seen from the edge ....

These Penrose ‘triangles’ are much better examples, as are the many other pictures available here: http://im-possible.info/

What do such pictures of ‘impossible objects’ tell us about vision — at least in (some? all? adult?) humans?

Caveat: they are not totally impossible!

Some pictures claimed to be of impossible objects do not depict truly impossible objects.



Bruno Ernst and Richard Gregory have shown that when looking at a ‘Penrose Triangle’ you may be looking at a real, totally consistent, 3-D object: they made examples!

Here is an example made by Bruno Ernst with its reflection in a mirror.

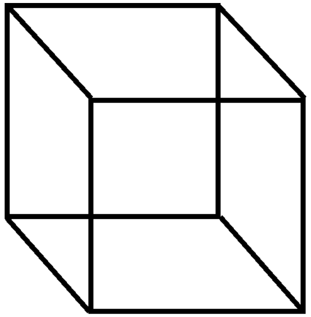
Notice that the mirror version shows us that the object visible in the foreground misleads us in more than one way. How?

The real 3-D ‘impossible’ objects are viewable only from a particular line of sight: Moving laterally makes the actual structure, which IS geometrically possible, visible.

From the ‘special’ viewpoint, the object is seen as impossible because local alignments drive a particular local interpretation for major picture fragments as all joined up.

In fact, not all are joined up – though their projections are – from one direction.

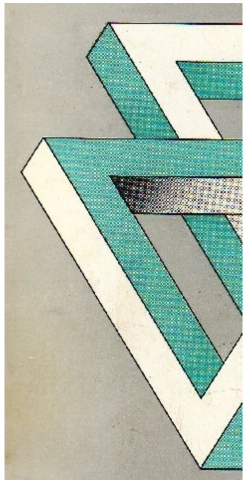
More can be worse

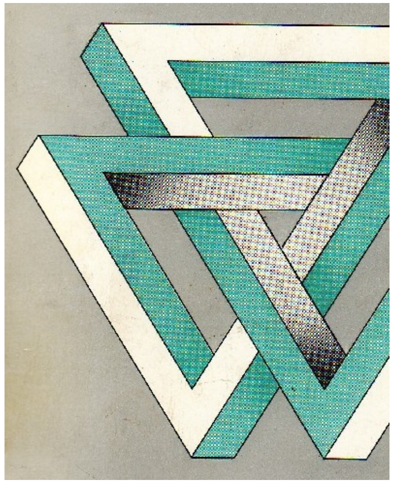


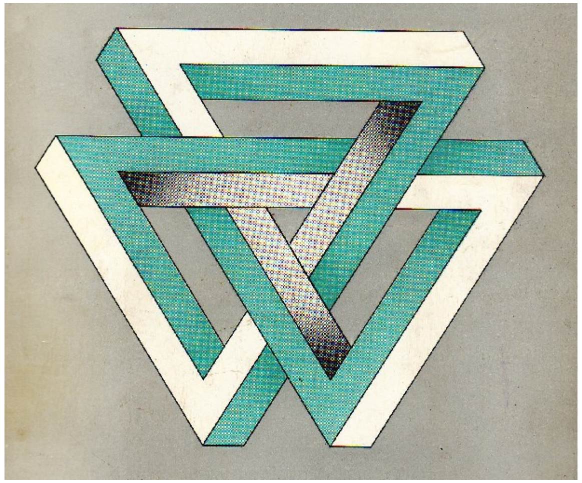
The ‘impossible object’ pictures depend on mechanisms that are part of everyday perception of structure.

A Penrose/Reutersvärd triangle is but one of many pictures that contain geometrically possible 2-D configurations of lines, which because of various strong cues drive our visual system to 3-D interpretation of various fragments in the picture, as does the Necker cube, whether static or rotating:

http://www.cs.bham.ac.uk/research/projects/cogaff/misc/nature-nurture-cube.html



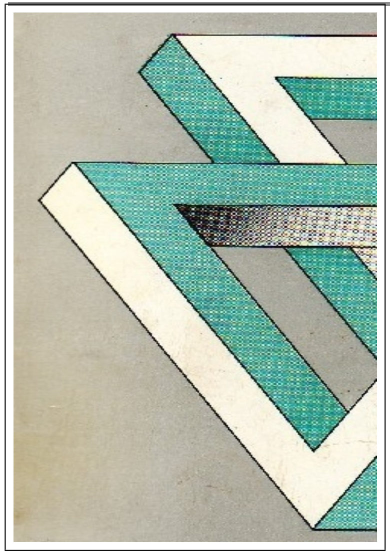




The ‘wire frame’ 3-D cube is perfectly geometrically possible. But in the Penrose, Reutersvärd and Escher pictures there is more local 2-D structure cueing specific 3-D structural interpretations, each internally consistent, but forming a globally inconsistent whole – e.g. violating transitivity anti-symmetry and irreflexivity of ‘further away’.

Sometimes various subsets (but not all) of such pictures are globally consistent.

Local structure in images and scenes



In the 1960s & 1970s, work by Guzman, Huffman, Clowes, Waltz, Turner, Barrow, Tennenbaum & others indicated how local image features combined with some prior assumptions about kinds of objects in the scene, could act as cues to 3-D scene fragments with specific geometrical structures.

Those prior assumptions might be false, but could be supported by more general features of the current view. E.g. if a scene appears to have many straight lines and especially if they remain straight across changes of viewpoint, the scene certainly includes many straight 3-D edges. Moreover, the surfaces joined by those edges must be planar (at least locally).

We ignore for now the fact that straight lines in 3-D space will not project into straight lines on a retina, leaving the task of identifying ‘projected’ straightness to be solved separately. (D. Philipona, J.K.ORegan, J.-P. Nadal, Is there anything out there? Inferring space from sensorimotor dependencies. *Neural Computation.*)

As the picture shows, local straightness and planarity in the scene generate very strong constraints regarding how different portions of edges, surfaces and objects are related in the scene.

Likewise when there is motion, assumed or inferred rigidity generates very strong constraints regarding both how things can change and how relatively remote scene fragments are related.

Barrow and Tenenbaum (1978) ‘Recovering intrinsic scene characteristics from images’, pointed out other forms of inference.

We don’t know how many such inferences human and other animal visual systems make but there seem to be many of them, some of which, though not all, depend on recognised objects.
E.g. Gregory’s ‘hollow mask’ demonstrates how strong evidence for concavity can be overridden.

But local consistency plus constraints can produce global inconsistency – possibly undetected.

Further conclusions about biological (human) vision

The considerations above seem to have some strong implications for the mechanisms underlying at least human vision – which are probably shared with some, but not all, other animal vision systems.

Where edges and surfaces intersect, or come into view, or go out of view (e.g. because of self-occlusion), fairly easily detectable features of the optic array (or the image ?) can provide strong clues regarding local 3-D surface structure.

The clues can be misleading products of 'accidental' alignments, but if they persist across movements or between binocular views they are VERY unlikely to be spurious.

E.g. the Ames room and the Ames chair work only from a particular viewpoint.
There are other cases, like Gregory's hollow mask, which looks convex, despite rotational and other information to the contrary, but that's because of the combined effects of a lot of powerful cues on a much larger scale than the image fragments discussed here.

As Barrow and Tennenbaum, Horn, Ullman, and others, showed there are also other sorts of clues to local 3D structure, involving curved edges, changes of shading etc. Effects of motion will be even stronger and we probably need a project to SYSTEMATICALLY investigate and categorise them.

These (often monocular) cues will provide information about local 3-D structure that is far more detailed and definite than anything that can be provided by stereo since they depend on principles of geometry rather than the behaviour of intrinsically unreliable sensors.

Humans have ways of combining such local structured 3-D interpretations into more complex interpretations, but because consistency checking is intractable, they can be 'fooled' by impossible objects - but only when depicted in 2-D pictures: 3-D configurations can be checked by motion.

Implications

One conclusion is that seeing a 3-D configuration does not involve constructing a model isomorphic with what is seen.

For instance, in the Escher 'waterfall' picture what is seen is internally inconsistent and therefore there cannot be a model isomorphic with what is seen.

The 2-D projection is NOT isomorphic with what it depicts. (Sloman 1971)

There are many locally consistent fragments: so you might think they are all represented by isomorphic models.

What is experienced is 'all joined up', and if the fragmentary interpretations were isomorphic with the 3-D structures depicted locally, they could not be joined up.

That implies that although the local 3-D structural information is perceived, and therefore represented, the form of representation used is at a level of abstraction that allows impossible things to be represented.

Compare:

Is this inconsistent?

A is bigger than B and B is bigger than A

What about this?

A is bigger than B, and B is bigger than C, and D is smaller than C and A is smaller than D

Those are easy to check, but in general, telling whether some complex structure is inconsistent or not requires space or time resources that are at least an exponential function of the number of components in the structure.

How are impossibilities detected?

As pointed out in several of my papers, presentations and ‘usenet’ postings: humans perform well on many small problems but we do not ‘scale up’ (as computers sometimes do), though we ‘scale out’, insofar as we are able to combine old competences in novel ways.

Whether an impossibility is detected depends on whether the visual architecture includes mechanism that can inspect and reason about what is represented (either automatically, driven by the data, or when a question arises).

Somehow in the Penrose triangle the detection of the inconsistency seems to be automatic (data-driven) in many adults, though not in more complex figures where detecting the impossibility implies longer chains of reasoning.

This is one among many examples of humans having an ability that does not ‘scale up’, though no doubt it grows and improves between childhood and adulthood.

Presumably toddlers do not see the ‘penrose impossibility’?

There may be older humans in whom it is entirely lacking – e.g. perhaps as a result of a genetic or trauma-induced brain deficit.

The requirement for consistency in what is experienced was discussed in a message to the Psyche-B mailing list in December 1997, disagreeing with Bernie Baars’ claim that conscious experience uses a ‘global workspace’ whose contents must be consistent.

http://www.cs.bham.ac.uk/research/projects/cogaff/misc/consciousness-postings/psyche-b-1997-12-8.html

Implications for representation of scenes

There are many implications of these examples, and I suspect there are more implications than anyone has noticed so far.

Representations of scene fragments may include complete and precise information, but the fragments are assembled in a way that does not include complete and precise information about their relationships – perhaps because that can usually be safely left to the environment to do???

The information about fragments is linked together in a way that is loosely in registration with the optic array (different bits of which are sampled at different times).

In small fragments of the scene, relative depth information is represented (i.e. which bits are nearer and which further, and which lines and surfaces are sloping away in which direction, etc., etc.).

In small local regions, depth changes may be represented absolutely relative to some intrinsic scene measure, but these depth measures are not consistent across the whole scene: transitivity does not propagate around all visible fragments.

However depth information is not represented in a precise way that is globally consistent across the whole scene.

NOTE: I need to try to work out whether or not this is consistent with the 'retinoid' theory of Arnold Trehub in *The Cognitive Brain*, MIT Press 1991. I suspect that theory is too committed to a kind of 3-D isomorphism that would not allow representation of globally impossible objects (like Escher's water tower).

I don't yet know whether these ideas will be readily expressed in terms of standard computer data-structures. But I have some ideas regarding how it might be done (arising out of discussions with Jeremy Wyatt, about further generalising generalised aspect graphs).

Implications for Labyrinthine Architectures

Note: added 11 Apr 2007

The discussion above is not meant to imply that there is one representation of the environment produced by or partially produced by the vision system.

- In 'On designing a visual system (Towards a Gibsonian computational model of vision)', `http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#7` in *Journal of Experimental and Theoretical AI,* 1989, I argued that instead of monolithic architectures for visual systems we need labyrinthine architectures, where different components of the visual system perform different sorts of tasks, with many connections to other parts of the whole system.

  The so-called 'ventral' and 'dorsal' visual pathways, not known to me at the time of writing, are but two among many conjectured in that paper.

- In particular, explicit, enduring, representations of information about the scene that are useful for planning actions, reasoning, describing, explaining are different from representations required for fine-grained control (visual servoing), most of which are not used consciously, and exist only transiently, whilst used. I.e. the latter are not explicitly accessible for multiple purposes over time.
  There are probably intermediate cases (e.g. short term proprioceptive memory?)

- Whether the kinds of inconsistency described here in the more explicit enduring representations can also occur in the latter transient implicit representations is an open question. (Jeroen Smeets claims the latter can also include inconsistencies. See Ref below.)

Various kinds of impossibility Added: 24 Apr 2007

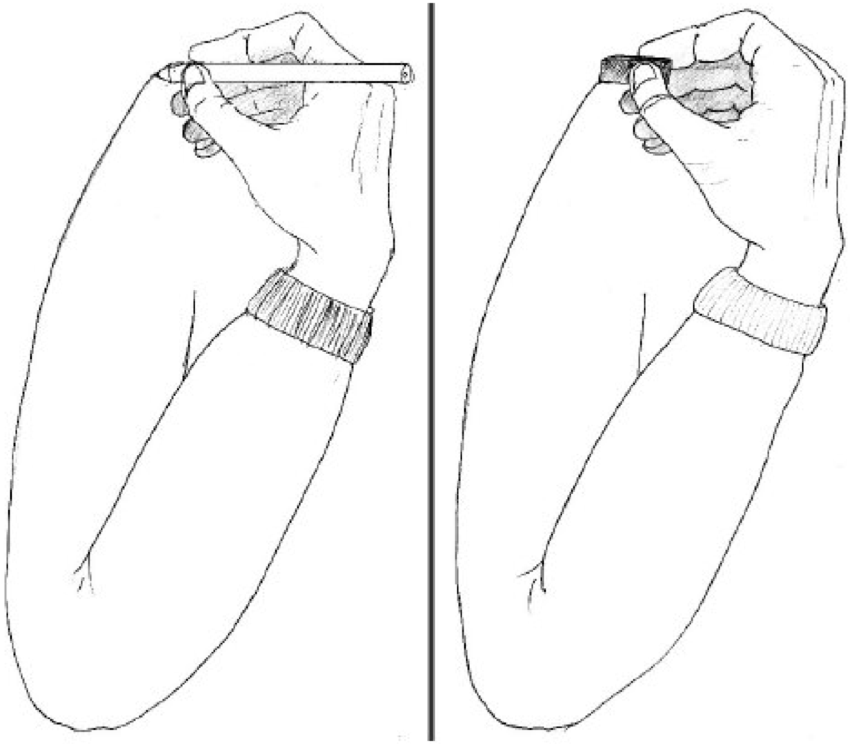
The examples above are all concerned with geometric constraints and geometric impossibilities. Some of Escher's pictures involve other sorts of impossibility, e.g. his picture of two hands, each of which is drawing the other. This seems to involve multiple impossibilities, not all of which are geometrical.

Manfred Kerber raised the question whether the impossible pictures have anything in common with logical paradoxes. This was not the main concern of this discussion, but it is worth considering.

Russell's paradox concerns the set R which contains all sets that do not contain themselves. If R contains R then it does not contain R, and if R does not contain R then R does contain R. In contrast there is no such contradiction in the notion of the set of all sets that do contain themselves.

Compare a hand drawing itself. It is impossible for such a process to get started. However, if a hand is erasing itself, then it is arguable that if it can finish then it cannot finish, and if it cannot finish then it can finish.

Though it has entertainment value I am not sure this sort of example gives us any new insight into what seeing is or how it works.



Draw Me Erase Me: Alison Sloman (after M.C.Escher)

Analogical and Fregean Representations

Added 25 Apr 2007

The Penrose triangle and some other impossible figures show that important qualifications are needed to the bit of Sloman 1971 which states that because analogical representations necessarily satisfy constraints imposed by the medium used they can reduce the search requirements for some problems.

> “....small changes in the representation (syntactic changes) are likely to correspond to small changes in what is represented (semantic changes); changes all in a certain direction or dimension in the representation correspond to similarly related changes in the configuration represented, and constraints in the problem situation .... are easily represented by constraints in the types of transformations applied to the representation, so that large numbers of impossible strategies don’t have to be explicitly considered, and rejected. Hence “search spaces” may be efficiently organised. By contrast, the sorts of changes which can be made to a Fregean, or other linguistic, description, ... are not so usefully related to changes in the structure of the configuration described.”

The pictures of impossible objects show, instead, that there are forms of representation that can adequately express what we see locally (including local spatial structures, relationships and affordances) in a way that rules out contradictions, while also allowing global inconsistencies that may be hard to to identify.

This rules out use of ‘models’ in the usual sense – i.e. representations that are isomorphic with the represented structures, since models cannot be inconsistent.

The virtues of line-labelling

Added 25 Apr 2007

The symbolic Huffman/Clowes vision systems of about 35 years ago were based on the use of constraints to eliminate local ambiguities, so in principle if there was a global inconsistency they would reject all local inconsistencies. But they could not detect the impossibility of a Penrose triangle because they did not represent enough metrical information.

The Huffman/Clowes line-labelling systems lacked the representational power to compute relative distances so the inconsistency in the Penrose triangle would not have been discovered. Does this make it a plausible part of a theory of how humans represent 3-D spatial structure, since they too don't easily discover all global inconsistencies?

Hinton pointed out in 1976 that the use of rigid local constraints meant that even a small amount of noise in an image could cause spreading 'gangrene' causing the whole image to be uninterpretable.

He proposed a 'relaxation' mechanism based on soft constraints to avoid that.

But the issue of coping with noise was orthogonal to the issue of representing local structure without expressing global structural relations.

References and thanks

I am very grateful to the Impossible Objects web site which provided several of the pictures:
http://im-possible.info/

Max Clowes first introduced me to the idea that the study of impossible and ambiguous figures could give deep insights into the nature of vision. http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-clowestribute.html

The ideas presented here arose partly from discussions with colleagues in the EU-funded CoSy robot project (including a conversation with Michael Zillich about limitations of current approaches to computer vision).
http://www.cs.bham.ac.uk/research/projects/cosy/PlayMate-start.html

Some of the ideas were also floating around in my email discussions with Arnold Trehub about what can and cannot be represented in the 'retinoid' model presented in his 1991 book *The Cognitive Brain*.
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/trehub-dialogue.html

In 1971 at IJCAI'02, I presented arguments that in addition to logical representations and modes of reasoning machines, like humans, would need spatially organised forms of representation. The paper is here
http://www.cs.bham.ac.uk/research/projects/cogaff/04.html#200407

A modified version became Chapter 7 of *The Computer Revolution in Philosophy* (1978)
http://www.cs.bham.ac.uk/research/projects/cogaff/crp/chap7.html

Chapter 9 of that book argued for vision as a multi-layer process assembling different sorts of information fragments at different levels http://www.cs.bham.ac.uk/research/projects/cogaff/crp/chap9.html

Work on the CoSy PlayMate robot led to the conclusion that vision was essentially about perception of processes going on concurrently at different levels of abstraction (in 2005):
http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0505

David Vernon pointed me at this slide presentation by Jeroen Smeets, on 'Why we dont mind to be inconsistent'
http://www.eucognition.org/embodying_cognition_2006/Jeroen_Smeets.pdf

Discussions with Manfred Kerber led to some of the later insertions. TO BE CONTINUED