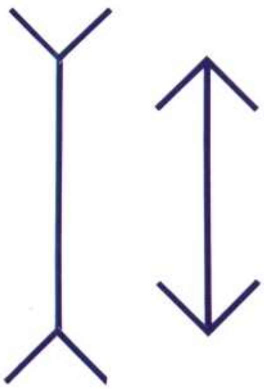
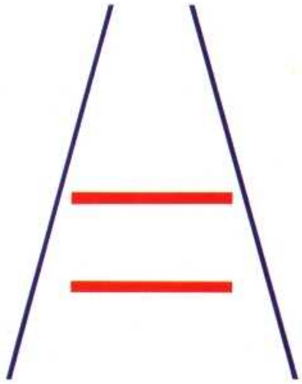
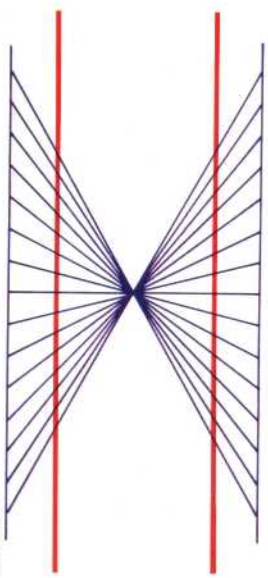
What You Think You See



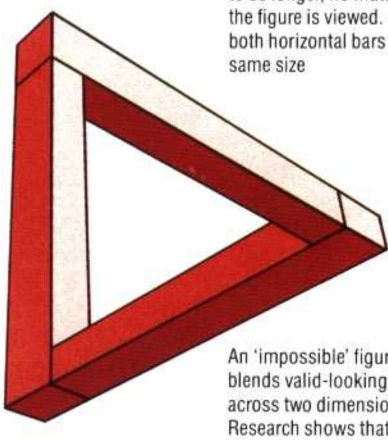
Which is longer? The Muller-Lyer illusion makes you think the vertical line in the left-hand figure is longer than the vertical line on the right. In fact, they are the same length



In the railway illusion, the upper horizontal bar appears to be longer, no matter how the figure is viewed. In fact, both horizontal bars are the same size

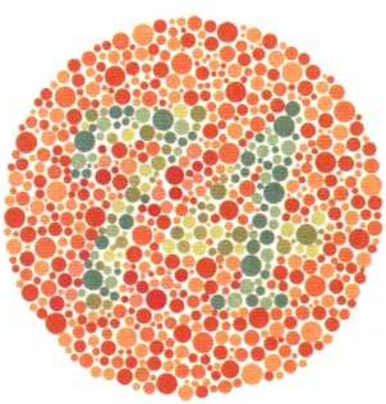


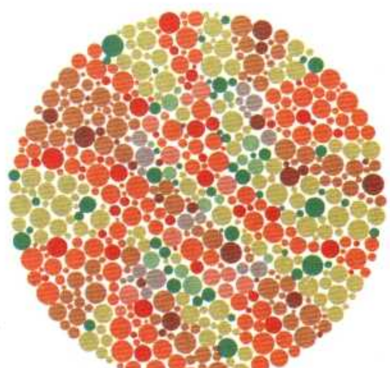
Although the vertical lines appear to bend, they are actually straight. The eye bends the image to conform with the spread of the rays



An 'impossible' figure, which blends valid-looking shapes across two dimensions. Research shows that humans are unable to recognise this as an object

KEVIN JONES





The colour patterns shown, taken from Ishihara's tests for colour blindness, measure red-green deficiencies. In the first example, persons with normal vision will see the number 74, while persons with a red-green deficiency will see the number 21. In the second example, normal vision will see nothing, or only a dim outline of a figure 2, while red-green deficient viewers will clearly see a 2

by the amount of light falling on that particular byte. These devices are becoming increasingly cheap and provide an area of RAM memory that contains all the information concerning the scene viewed by the robot's eye.

In general, the output from a robot eye is held in a two-dimensional array, each element of which contains a value that corresponds to the brightness of the light falling on that particular part of the scene being viewed. The number of elements in the array gives the *resolution* of the image, and the range of numbers that can be held in each array element determines the number of *grey scale* levels that can be discerned. Traditionally, in vision systems each element of the array is called a pixel, or 'picture element'. So an image array of 500 by 250 pixels that represents brightness levels by allocating one byte to each pixel would have a horizontal resolution of 500 pixels, a vertical resolution of 250 pixels, a total of 125,000 pixels, and 256 grey scales ranging from black to pure white. To give an idea of the detail that such a picture might provide, consider a standard television picture. The British system uses 625 vertical lines, so the vertical resolution is 625 pixels. To get a similar resolution across the screen, approximately 1,000 pixels would be needed (because the screen is wider than it is high), and the grey scale levels could be represented by the same single byte to give 256 brightness levels. A robot system with an equivalent resolution would give an acceptable image for the computer to process.

SCANNING THE IMAGE

The processes that must be undertaken by the robot 'brain' in order for this image to be 'seen' follow a set pattern. The first step involves adjusting the grey scale levels so that adjacent pixels with similar grey scale levels are 'smoothed out' to the same level. The computer works over the entire image area averaging levels so that any small irregularities are removed. Once this is done, the computer examines the image again, noting any adjacent pixels that have markedly different grey scale levels; these differences are then emphasised. The idea behind this is that important features in the image are probably marked by boundaries, such as lines and edges, which will show up as sudden changes in the grey scale level: the computer takes note of these, emphasising them to make sure that they stand out.

Once this has been done, the computer scans the image again, searching for all the really large changes in grey scale levels. It then uses these in much the same way as a human would join the dots in a puzzle to make up a picture. In most of these puzzles, the human is aided by the fact that the dots are numbered; the computer has no such help and must simply follow what seems to be a likely route. At the end of this process the robot will have an internal picture of the scene before it in which it has 'smoothed' the image out and drawn lines around those objects that seem to be important.

But is this 'seeing'? In fact, all the robot has done is to carry out certain transformations of the scene — in other words it still does not 'know' what it is looking at.

There are two solutions to this problem. The first is to program the robot with a set of rules that are expressed as a set of simple statements about the visual world. This is known as a 'bottom-up' approach to visual perception — so called because the robot starts with very simple things and tries to work out from them what it is seeing at a more complex level of understanding. The second approach is to program the robot with a set of