The human visual system

Vision and hearing are the two most important means by which humans perceive the outside world.

1 Low-level vision

Light is the electromagnetic radiation that stimulates our visual response. It is expressed as a spectral energy distribution $L(\lambda)$, where λ is the wavelength that lies in the visible region, 350nm to 780nm. Light received from an object can be written as

$$I(\lambda) = \rho(\lambda)L(\lambda)$$

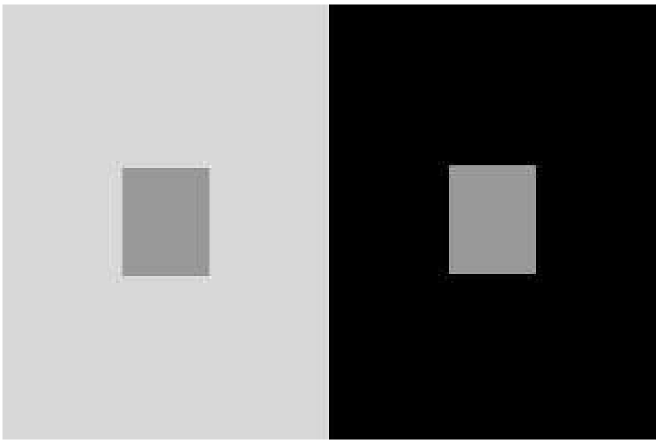
where $\rho(\lambda)$ represents the reflectivity or transmissivity of the object and $L(\lambda)$ is the incident energy distribution. The illumination range over which the human visual system can operate is roughly 1 to 10^{10}, or ten orders of magnitude.

The retina of the human eye contains about 100 million **rods** and 6.5 million **cones**. The rods are sensitive, and provide vision the lower several orders of magnitude of illumination. The cones are less sensitive, and provide the visual response at the higher 5 to 6 orders of magnitude of illumination.

The cones are also responsible for colour vision. They are densely packed in the centre of the retina (the fovea), at a density of approximately 120 cones per degree of arc subtended in the field of vision. The density of cones falls off rapidly outside a circle of 1 degree radius from the fovea.

The pupil of the eye acts is an aperture, and therefore acts as a lowpass filter. In bright light the passband is about 60 cycles per degree.

The two smaller squares in the figure below have equal luminances, but the one on the right appears brighter:



The reason is that our perception is sensitive to luminance *contrast*, rather than to absolute luminance.

According to Weber's law, if the luminance f_0 of an object is just noticeably different from the luminance f_s of its surround, then their ratio is

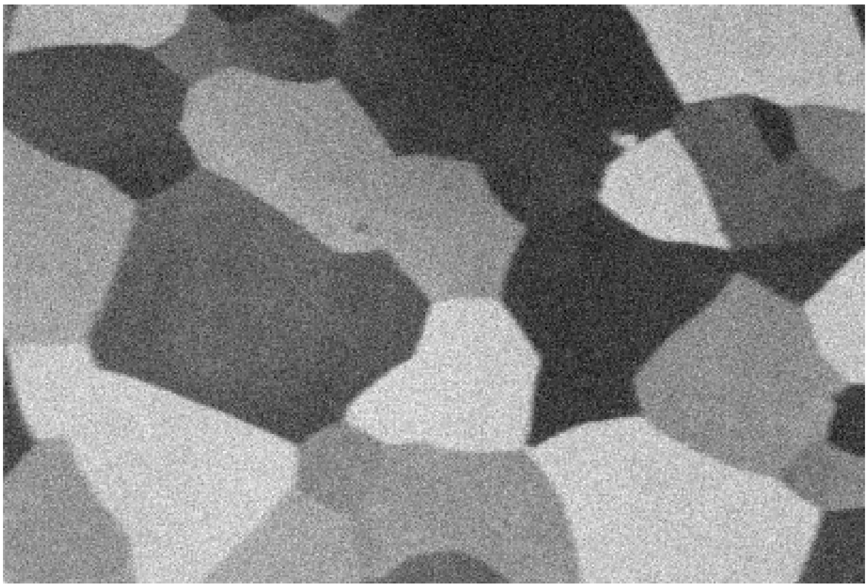
$$\frac{|f_s - f_0|}{f_0} = \text{constant}.$$

Writing $f_0 = f$ and $f_s = f + \Delta f$, this can be written as

$$\frac{\Delta f}{f} \approx c.$$

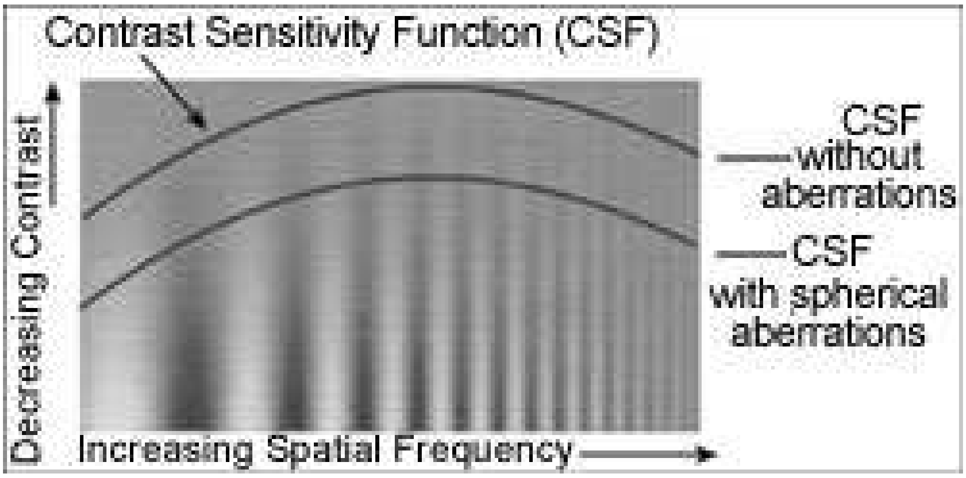
The value of the constant has been found to be 0.02. This equation states that equal increments in the *log* of luminance should be perceived to be equally different.

This model partly explains why a uniform level of random noise is more visible in a darker region than in a bright region. Consider the image below, which is the result of adding zero-mean white noise to an original undegraded image:

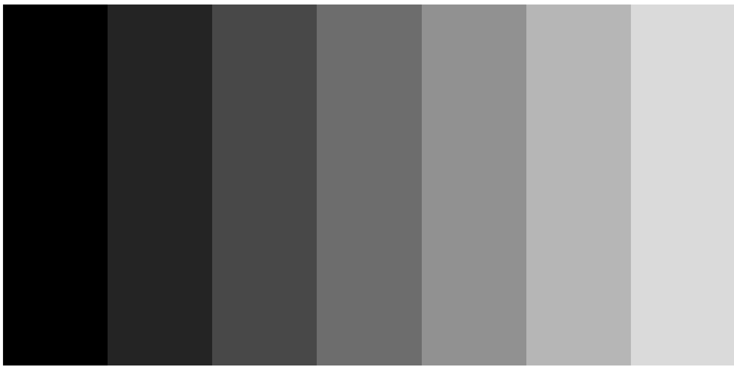


The grainy appearance due to noise is more pronounced in the darker uniform regions than in lighter regions.

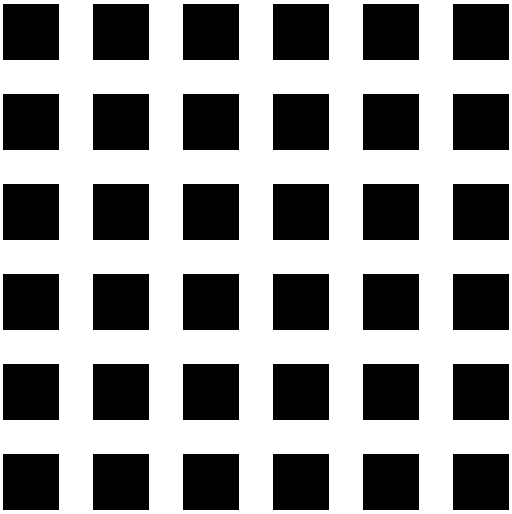
Apart from our perception of intensity being a nonlinear function of actual intensity, the human visual system also has varying sensitivity to different spatial frequencies:



The spatial interaction of luminances from an object and its surrounds also creates a phenomenon called the **Mach band** effect, which shows that brightness is not a monotonic function of luminance. In the bar chart below, each bar has constant intensity:

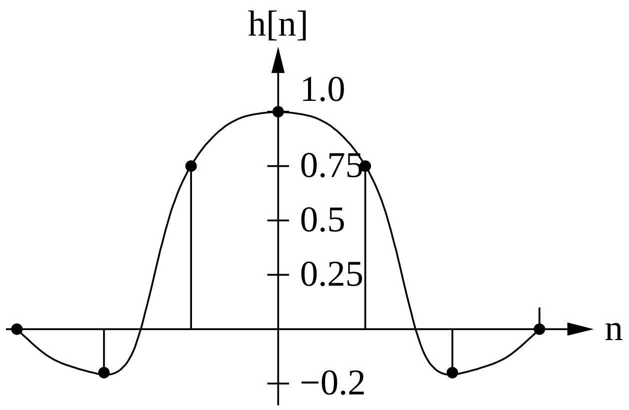


However, for a given bar the region looks brighter towards the left and darker towards the right. A similar effect is observed in the following image:

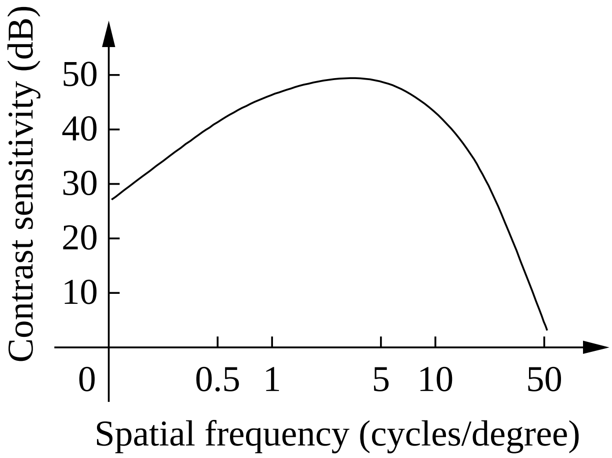


Although we know that it is not the case, the white areas between the black squares appear with varying shades of grey.

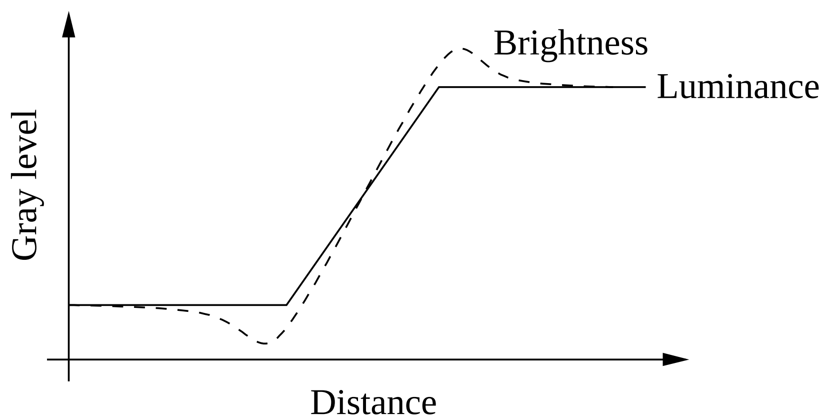
Both of these observations can be explained by the fact that the visual system applies spatial filtering to the signal, and that the impulse response exhibits significant overshoot and undershoot.



The visual system therefore exhibits a bandpass nature, with frequency response



This is partly responsible for uneven brightness perception within a region of uniform intensity when there are edges present. In the case of Mach bands, for example, the visual system perceives the step edges according to the response

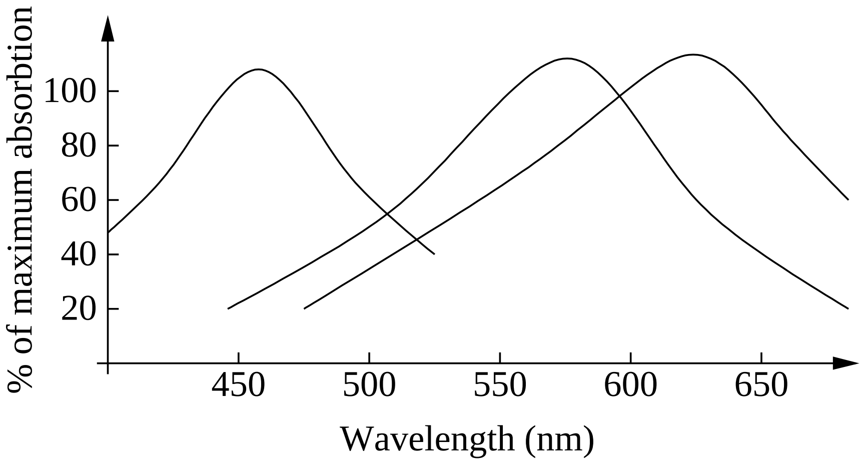


An effective model of the HVS must take into account all (and more) of the factors described, and can clearly become quite complex.

Aside from the response to grey levels, of which we can observe a few dozen, we have the ability to distinguish between thousands of colours. The perceptual attributes of colour are *brightness*, *hue*, and *saturation*:

- Brightness represents the perceived luminance, as mentioned before
- The hue of a colour refers to its “redness”, “greenness”, and so on. For monochromatic light, differences in hue are manifested by the differences in wavelength
- Saturation is that aspect of perception that varies as more white light is added to monochromatic light.

Colours are sensed by the eye by three different types of cones, which are sensitive to different wavelengths. Their spectral responses peak in the yellow-green, green, and blue regions of the visible electromagnetic spectrum:



It is therefore possible in our *perception* of light to describe any colour by three values. Note however that the light itself forms a continuous spectrum of wavelengths.

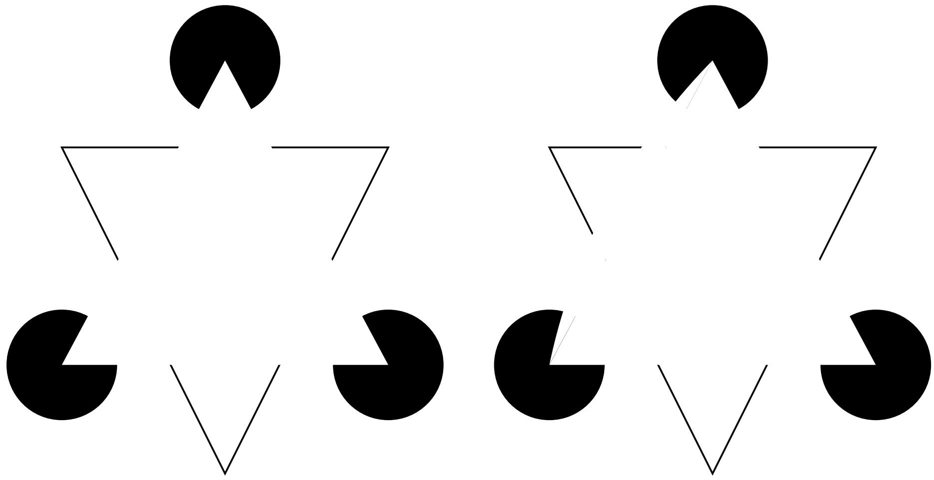
Colour spaces are a way of organising the colours perceived by human beings. The RGB colour space expresses a colour by its red, green, and blue components, so any colour can be represented by a point within a 3-D cube. However, although the RGB space is good for acquisition or display of colour information, it is not particularly good for explaining the perception of colours. Alternative colour spaces more suited to this purpose are HSV (hue, saturation, value), YUV (chrominance, luminance), and Lab, although numerous others exist. It is generally possible to convert from one colour space to another by applying a nonlinear transformation to the colour values.

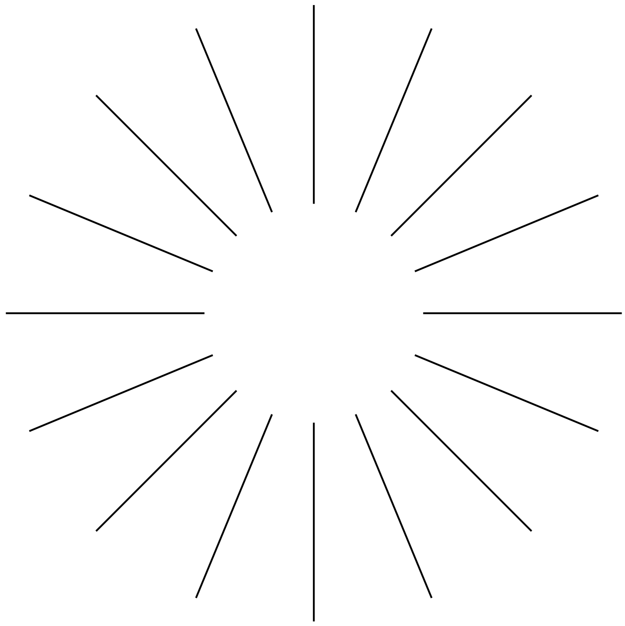
The theory of colour vision is vast, and a multitude of models exist for attempting to describe it.

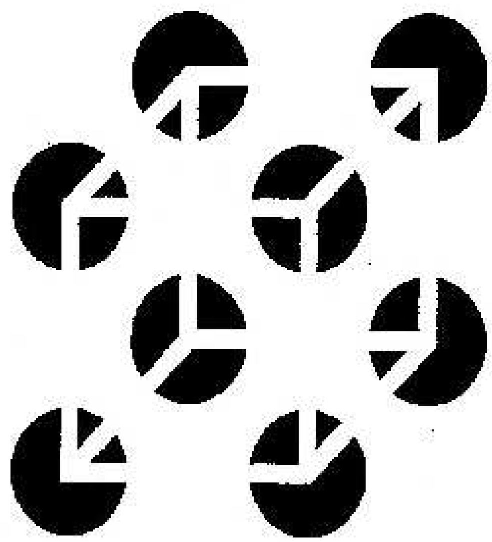
2 High-level vision

Aside from the low-level "hardware" aspects of the imaging process, the human visual system exhibits a considerable cognitive component. Vision is therefore also influenced by memory, context, and intention. There is evidence that these high-level components feed back to the low-level vision system, further complicating the issue.

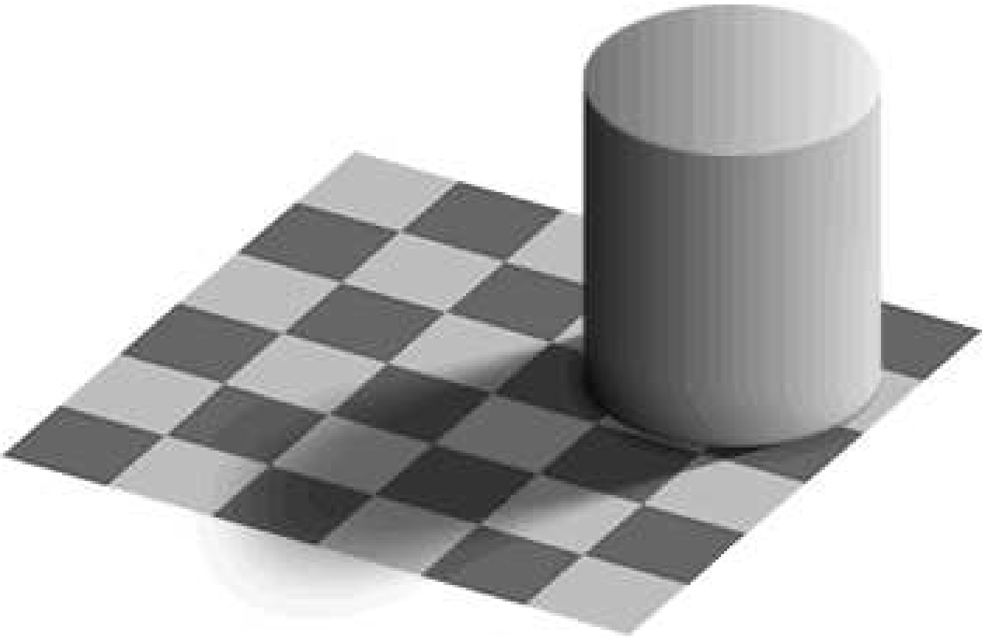
Subjective contours are an interesting aspect of the human visual system:

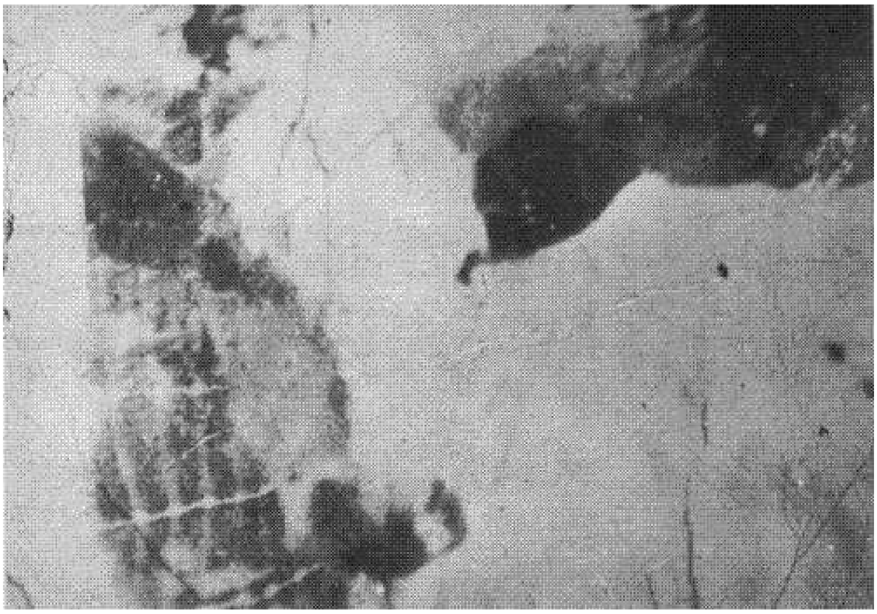






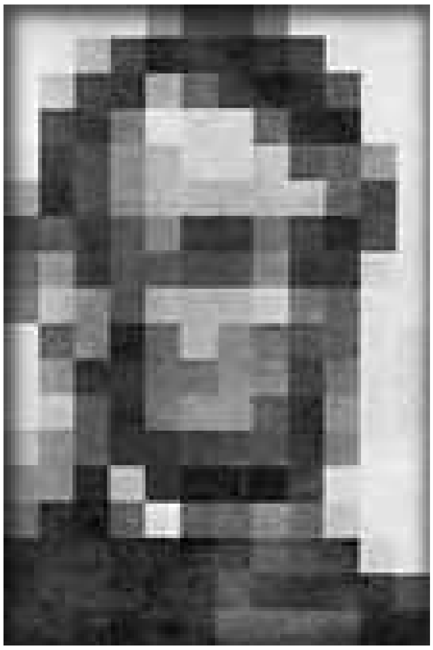
It is also apparent that a considerable component of the human visual system relies on us having a vast sum of knowledge regarding the appearance of objects in our environment:





We know that edges are very important in vision, but...

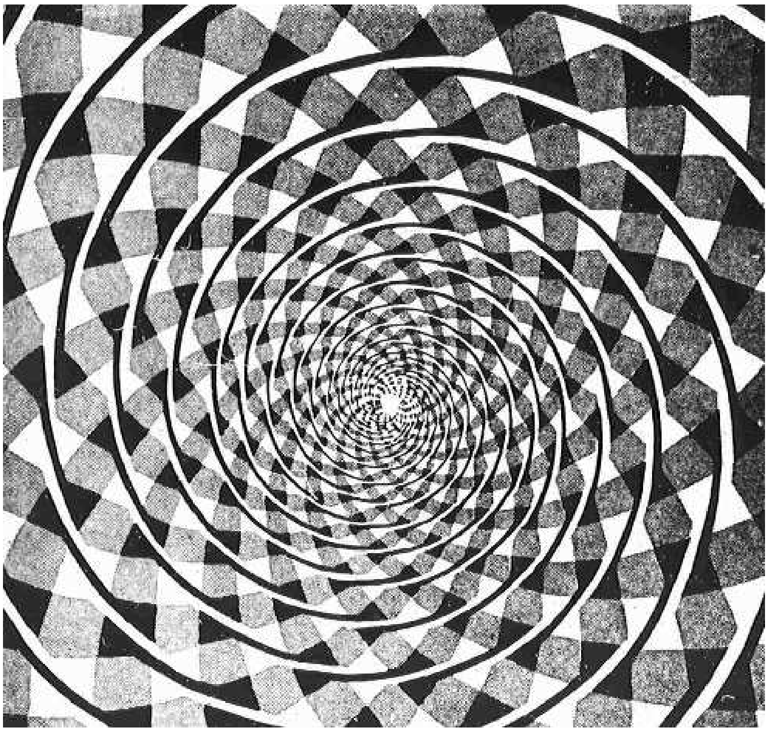
We’re very good at recognising faces:



We’re also quite good at recognising objects, such as chairs


And characters.

A complete theory of vision must be able to explain all of the “strange” illusions out there:



Finally, what's this all about?