Overview of the module so far
   Search
   Logic
   Planning
   Knowledge Representation
   Biological Intelligence

Overview of the module to go
   Vision
   Machine Learning
   Overview
   Revision
   Exam!
Today’s lecture

Vision is a difficult problem
The world as illusion/simulation versus direct vision
Building up complex representations from simple components
Object recognition
Vision is a difficult problem

“Vision is easy, it operates just like a TV camera, taking in images and projecting them in the back of the brain!”

But who or what does the viewing of images in the back of the brain?

There is no homunculus to do this inner viewing, (and if there was we would still have the problem of explaining how the vision system for the homunculus worked).
A key problem that vision, and other perceptual systems have to overcome is that the sensory systems produce a large amount of parallel data.

What specific aspects of this data should be attended to?

Perception and attention are linked to issues of self-awareness and consciousness (though the terms ‘self-awareness’ and ‘consciousness’ really cover several different kinds of phenomena)

Theories of perception usually suggest some form of non-attentive parallel processing occurring before a narrower, attentive form of perception.
Examples of world as illusion/simulation

Our conscious visual experience is of a seemless world.

However, this world is actually constructed for many snapshots (saccades) of our environment.

The visual system puts this together into coherent perception.

Many visual illusions rely upon the simulation our mind is creating having two mutually exclusive ‘answers’
Examples of world as illusion/simulation

Sometimes simulations created by our mind are based upon partial information, that is seen to be false when the overall picture is seen.
How many legs does this elephant have?
Examples of world as illusion/simulation

Our visual system uses 2D cues to build up 3D simulations, that is we infer 3D structures from 2D evidence
Object recognition in nature

How are the point images provided by all the nerve cells at the back of the retina processed so that complex shapes can be represented?

Simple mechanisms of recognition - Tinbergen (1951) - sticklebacks ignore very realistic models that don’t have red underbellies, and recognise very unrealistic models that do have red underbellies.
Building up shapes from simple cells

Hubel and Wiesel (1962) - low level feature detectors

On-off cell

Off-on cell
Building up shapes from simple cells - bar detector complex cells
Building up shapes from simple cells - edge detector complex cells
Hyper-columns in the visual cortex

Each column represents a bar or edge at a slightly different rotational orientation
Pandemonium: an AI application that possesses a hierarchy of increasingly complex features
Can hyper-columns explain how complex objects are perceived by natural organisms (such as humans)?

**NO**, simple features are NOT built up into ever more complex and abstract features. This is **not** how it is believed that object recognition occurs in humans. Single neurons **don’t** represent abstract objects (like your Grandmother or yellow volkswagens).
Object recognition

Objects are made up of lots of bars, edges and other components, but complex objects are not recognised as more abstract feature sets or templates.

Features and abstract templates fail to capture some structural relations (e.g. orientation of template) and discard all information that might be used to discriminate different instances of the same pattern.

Current theories of object recognition suggest that objects are represented with structural descriptions (that are symbolic though non-linguistic) propositions

(e.g. a ‘T’ is a vertical and horizontal line where the vertical line bisects and supports the horizontal)
Marr and Nishihara (1978) suggested that objects are recognized by forming 3D representations composed of cylinders and cones.
Different objects are formed by different types of cones, or cones in different arrangements or different relative proportions.

For example, a human is distinguished from an ape because it has relatively shorter arm-cones and relatively longer leg cones.

Objects are recognised from 3D representations, via 2½D intermediate representations.
Biederman (1987) developed these ideas. He extended the idea of cones to geons (geometric ions - which are for images what phonemes are for language). According to Biederman, objects can be recognised directly from 2D images because of nonaccidental properties of edges in images which can be used as reliable cues to related properties of edges in the world (for example, straight lines in images reflect straight lines in the world, and not ‘accidental’ straight lines)
Problems with Biederman’s (1987) theory

Geons may be used to distinguish faces from telephones, cats or washing machines,

But what kind of representations are used to distinguish one face from another (or one cat from another, or one telephone from another etc) when objects of a similar type may be built up from similar sets of geons.

Tarr and Pinker (1989), and others, suggest that geons are not the only representation used in object recognition. This is because geons are view invariant representations, and there is lots of evidence that object recognition is not always a view invariant process
Conclusion

Vision involves forming internal representations of some sort (simulations/emulations)

Low level aspects are represented by features, but there is no evidence for single neurons that act as abstract feature detectors.

Pandemonium is an AI application that can recognise visual objects like letters, considering them as high level sets of features, but this approach is less flexible than human object recognition.

In modern object recognition theories, higher level aspects are more likely to be represented with some kind of structural description as proposition (though these may be non-linguistic).

Different kinds of representations will be used for different kinds of problem,