Introduction

An outline process for object recognition:

(i) Requires processes to construct a description of the object perceived, based on the retinal image.

(ii) Requires processes that can store the description so we can recognise something again.

(iii) Requires comparison processes - so we can compare what we see with what we've stored previously.

(iv) Need processes to enable us to recognise the same object from different angles. The nature of this mechanism is important and controversial.

Humphreys and Bruce (1989) - the way object recognition fits into the wider context of cognition - including perception, categorisation and naming.

Early visual processing

|

Viewpoint dependent object descriptions

|

Perceptual classification

|

Semantic classification

|

Naming

Chapter4 - Recognition

Different types of recognition

Object and face recognition - **Humphreys and Bruce** model ends at naming - but not necessary for recognition (animals recognise objects), but humans tend to distinguish between 'an apple' and 'Sigmund Freud' - one being a class of objects, the other a specific instance. (between-category vs within category distinctions.)

Therefore, face recognition is often researched differently to object recognition, plus internal features of faces can move, movement expresses social and emotional cues and faces change over time - e.g. ageing.

Different types of face recognition occur:

(i) Between familiar/unfamiliar faces. **Pike et al** - E-FIT images can be recognised even if rated as poor likenesses by other participants - but images used were of famous people. **Kemp et al** - witness identification of a suspect not good - even if the anxiety of an identification parade is reduced by using video.

(ii) Recognising what emotion a face may be conveying can be performed with high accuracy. **Young et al** - evidence that while we do have specific processes for recognising emotions, these aren't used in recognising identity (makes sense we need to be able to tell if someone is angry even if we don't recognise them.)

No definitive answer to whether the processes used to recognise objects are the same as for faces, but they are usually treated as different research areas.

Active processing - recognising objects by touch. A limitation of the **Humphreys and Bruce** model is that it is passive. **Gibson** - stresses perception is an active process - and so recognition is likely to be active too - particularly if the sense of touch is considered.

We have great control over our hands. Fingers are moved precisely and we can vary the pressure we exert on objects through the use of **feedback**. **Kinesthesis** - the knowledge of where our limbs are - is combined with this. Processes that allow us to keep track of where our limbs are relative to each other are known as **proprioception**. All these create **haptic information** - which can be used to generate an object description.

Lederman and Klatzky - found consistency in the way people use their hands to obtain haptic information, by using **exploratory procedures**. Their later research provided information on what procedures were used - e.g. if the texture of an object is important to its recognition, then people move their fingers over its surface.

Haptic information provides better knowledge about the weight and textures of objects; visual information is better at providing 3D shape information. Haptic perception is (always?) active - therefore recognition is not necessarily passive as implied by the **Humphreys and Bruce** model.

Recognising 2D objects. Can also distinguish between type of recognition if an object is 2D or 3D in nature. Early research focussed on using 2D images and patterns, which arguably tell us little about 3D object recognition.

Simplest model of visual pattern recognition is **template matching** - the idea we have a large number of templates in long term memory for shapes of letters, number etc. This concept fails to deal with the huge variations in patterns for even alphanumeric characters, however.

Therefore, some way of accounting for variations is needed. **Feature recognition** is one approach - e.g. an 'O' is a single continuous curve; 'P' is a vertical line and a discontinuous curve. Pandemonium system (**Selfridge** - Morse code; **Neisser** - alphanumerics) is an example - but these theories do not capture structural relationships between features e.g. '^' could be misrecognised as a letter 'v'.

A more successful approach is one based on **structural descriptions** - a set of propositions (expressed in language or symbolic representations) that describe the elements that make up and object and the relationship between those elements.

This method could also be applied to 3D objects - but it also means we need to be able to turn our 2D retinal images into 3D descriptions that are object, rather than viewer centred. These processes are the second part of **Marr**'s theory of vision.

Object centred vs viewer centred descriptions - a simple pattern matching program would mistake a coffee cup shown from a different angle to its own internal representation as not being a coffee cup - a viewer centred description.

Recognition therefore needs to take place independently of the viewpoint of the observer.

Marr - conceptualises this are turning the viewer-centred 2½D sketch into a 3D object centred description, allowing recognition to take place from any angle.

Recognising 3D objects

The second part of **Marr**'s theory concerns itself with how a 2½D sketch (viewer centred) is turned into a 3D (object centred) description. **Marr and Nishihara** proposed the use of a **canonical coordinate frame** - each object is represented in a framework that approximates to the shape of the object.

First step in establishing the canonical coordinate frame is to define a central axis for the object. This is so important a step it is restricted to objects that can be descried by one or more **generalised cones** - many natural objects (tress, animals) can be represented this way, but nonetheless it is a weakness of the theory.

To do this, the information in the 2½D sketch is used to work out the shape of the object, based on its **occluding contours** (the object's silhouette). Points on the surface that = boundary of the silhouette are particularly important - they are referred to as the **contour generator**, as this is what they can be used to do.

However, the same (2D) silhouette can be produced by different (3D) objects. **Marr** suggests this problem is resolved by the visual system making assumptions about what we are seeing - we can interpret silhouettes because there is a source of additional information that constrains how we perceive them.

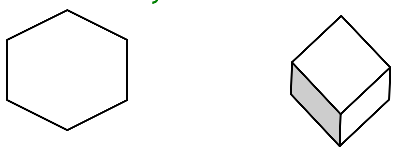
Marr suggests this additional information comes from three built in computational assumptions:

(i) Each point on the contour generator = a different point on the object

(ii) Any two points close together on the contour in an image are also close together on the contour generator of the object

(iii) All the points on the contour generator lie in a single plane.

Point (iii) is vital to the theory, but is problematic. E.g. the silhouette of a cube may produce a regular hexagon. Marr's assumption (iii) is violated as the point on the cube that produces the top of the hexagon is further away than point that produces the bottom of the hexagon - so the silhouette is not interpreted correctly.



Next step is to find the axis/axes necessary to represent the object - straightforward for simple shapes; and complex shapes are broken down into components (**primitives**), with a single axis per component.

One method for locating axes - (i) work out areas of **concavity** and **convexity.** (ii) divide shape into sections by using areas of sharp concavity. (iii) use these to divide object into smaller parts. (iv) each section then has a component axis, which can then be represented in relation to the horizontal axis of the body.

Once a 3D description is created, **Marr and Nishihara** argue the next step in recognition is to compare this to a catalogue of 3D models - generated by 3D descriptions of all previously seen objects. Catalogue is hierarchical, highest level is descriptions of objects not decomposed into components, next level has more detail and so on.

The 3D model to the new object or target is related to the catalogue at the highest level. Target is then compared at the next levels until a match is found and the object recognised.

Evaluation of Marr and Nishihara's theory

The key prediction of the need to establish a central axis to aid recognition is supported in a study by **Lawson and Humphreys**. Recognition of rotated objects did not have an effect unless the major axis was tilted towards the observer, when recognition was disrupted - presumably as the foreshortening of the axis made it harder to locate.

Warrington and Taylor - neuropsychological case study. Patients with damage to an area of the right hemisphere could recognise objects in a typical view but not when presented in an atypical manner. Also found it difficult to say if an object was the same when presented typically and atypically simultaneously as a photograph. This could be explained as the patients being unable to transform a 2D version of the atypical view into a 3D model description.

However, it could be that as well as the axis being difficult to find, rotation causes key object features to be hidden. A later study by **Humphreys and Riddoch** controlled for this possibility by using images where either a critical feature was hidden or the axis had been foreshortened through rotation. More problems were found with axis foreshortened than hidden feature images, providing some support for **Marr and Nishihara**'s theory that **axis location plays a key role in generating 3D descriptions**.

Biederman's theory

Extends **Marr and Nishihara**'s work by not restricting component primitives to generalised cones. Instead, used **geons** - a basic set of cylinders and cubes - many of which are generalised cones but other 3D shapes are included in his set of 36.

The main difference comes from his argument that to generate a 3D shape **Marr**'s contour generators are not necessary. This is because each geon has a key feature that does not vary across different viewpoints. Therefore, all that is required is for features on the 2½D sketch to be matched to a geon so that a 3D structural description of the object can be generated.

Some regular aspects of 3D shapes remain constant in any 2D image of the object - **Biederman** calls these 'nonaccidental' properties. He lists five of these:

Curvilinearity; Parallelism; Cotermination; Symmetry; Co linearity

Choosing the right geon to represent a (part) object is about detecting these properties and selecting a geon that shares them. A 2D image of a ball is a circle and therefore contains curvilinearity and symmetry only. The only geon that matches is a sphere.

However, these assumptions can also lead to misinterpretation of an image. An edge on view of a wheel looks as if it has co linearity, as it appears to have two vertical edges. But this is not the case - it is only the viewpoint that makes it appear so.

There is however evidence to support **Biederman**'s theory.

Biederman - concavities are used to divide objects into components tested by giving participants images with that part of the contour missing. Greater disruption to recognition observed than removing part of the contour from elsewhere.

Biederman and Gerhardstein - investigated the extent to which recognition requires the generation of an object centred description rather than relying on the viewer centred description alone. Repetition priming used to see if presenting one viewpoint of an object would help it to be recognised from another viewpoint. Results showed priming occurred if change of angle was < 135 degrees and if one or more geons hidden between the first and second view, priming was less effective even if change in angle was < 135 degrees.

Supports the idea that an object centred description is generated **and** that it makes use of geons.

And evidence to suggest problems:

Bulthoff and Edelmann - participants unable to recognise complex objects from a novel viewpoint even if the view of the object was one that should have facilitated the creation of an object centred description. So it is unlikely that recognition is solely reliant on the generation of such descriptions. **Tarr** suggests there are tasks that may involve viewpoint-dependent descriptions.

It is hard for both **Marr / Biederman**'s theories to incorporate within-category discrimination, as a lot of information is lost in their hypothesised processes.

Both these theories are passive in nature - based on a single snapshot of a 2D retinal image. More active approaches to recognition can (and need to) be taken.

Face recognition

Problematic for the 3D model approaches so far discussed. These concentrate on the 'perceptual classification' stage of the model (**Humphrey and Bruce**). For a face, you need to know whose face it is. The within-category judgements required make it different from general object recognition. **Tanaka** argues face recognition is similar to expert recognition (e.g. bird-watching) - but it is something we all share and acquire without specific training. An innate processing system or learned skill?

Recognising familiar and unfamiliar faces

Bahrick et al - recognition tests reveal little forgetting of names and faces of school friends over a period of 35 years. Not the same for all faces - study of teachers found face recognition high for recent past students (69%) but fell to 26% after 8 years.

Yin - recognition of faces seen only once is good if tested immediately (93% recognition). **Bruce** found recognition rates dropped in expression or viewpoint was different - so perhaps Yin's results were about recognition of a picture of a face, rather than pure face recognition.

Kemp et al - matching unfamiliar faces shown together (no memory involved) is difficult. Cashiers and photo credit cards - accepted cards with a resemblance to a shopper (correct decision to reject occurred just 36% of the time) - and if a poor resemblance reject rate was 66%.

Bruce et al - similar findings for matching two high quality photographic images when face is unfamiliar. Participants shown a video still of a target; then presented in a line up of 9 similar images of others. When told target present, success rate of only 80%.

If not told or if pose of target varied between initial presentation and test, performance becomes much worse. **Burton et al** - some automated face recognition systems perform better at this task.

Kilgour and Lederman - participants explored faces visual and tactually. Performance no better than when faces explored by touch alone.

Modelling in face recognition

Young - diary study of 22 on errors made in recognising people. Errors made fell into a number of categories:

- Person misidentified - someone unfamiliar misidentified as someone familiar
- Person unrecognised - v.v.
- Person seeming familiar only (no further details retrieved from memory)
- Difficulty in retrieving full details (some semantic information, but not all retrieved)
- Decision problems - think you see someone you know but decide it can't be them for whatever reason.

=> before semantic information is retrieved, we need to realise that a face is familiar.

1980's - **Hay and Young, Young et al, Bruce and Young** - developed a cognitive theoretical framework - a sequence of stages for recognition.

On meeting, encode face; activates face recognition units (FRUs). If reasonable match FRU activated; access to semantic information (their identity, e.g. occupation) stored in PIN (person identity node). Once a PIN for a face activated; name generated.

A cognitive system is involved as well - information provided by the recognition system requires evaluating. Can be used to explain recognition problems - if we know someone lives a long way away yet we see them when we're not expecting to, our knowledge can override what we've seen - "decision problem".

Bruce and Young model has separate rotes for facial expression analysis, facial speech analysis and face recognition.

Sequential access of these different types of information is supported from lab experiments.

Hay - showing famous and unfamiliar faces. Names not retrieved without also being able to state the person's occupation; supports idea 'person identity' information is retrieved before the person's name.

Johnston and Bruce - faces classed as familiar more quickly than they can be classified by occupation; and classification by name takes longer than by occupation. Support for idea that perceptual classification takes place before semantic classification and that name is retrieved last.

Connectionist model of face recognition - IAC

Interactive Activation and Competition network

Burton et al (1991); Burton and Bruce (1993)

Computer simulation - tested by seeing if it is compatible with the evidence and by looking at the predictions it generates.

Made up of units, organised into four pools, containing:

(i) FRUs - one FRU per familiar person - recognition activates the appropriate FRU. Allow perceptual information to be mapped to stored memories (as per **Bruce and Young** model.)

(ii) PINs - face belongs to a person; one unit per known person

(iii) SIUs (semantic information units) - e.g. occupation

(iv) Lexical output - units representing output as words or name.

A route on word recognition (WRUs) is also present. WRUs have direct links to NRUs (name recognition). WRUs that are not names are linked to SIUs.

Many SIUs are shared - e.g. 'British' 'teacher'.

Face recognised by:

(i) FRU activated - increases activation in relevant PIN

(ii) PINs linked to SIUs - so PIN activation increased activation in relevant SIUs

(iii) Threshold reached on PIN = familiarity. Different types of information come together to do this - so familiarity is based on the result of pooled information.

Links can be excitatory or inhibitory - SIU for 'Mick Jagger' will excite many others - so more than one PIN will be excited in turn - e.g. for other singers. Model therefore incorporates priming effects - quicker to recognise Bill Wyman if you've already seen Mick Jagger.

Strength of the model is that it can account for findings from both lab studies and the everyday errors described in **Young**'s diary study.

Neuropsychological evidence

From studies of prosopagnosia (inability to recognise faces) - two key findings:

(i) *Identification of expression is separate to face identification*

(ii) *Face recognition and awareness of face recognition may also be independent of each other.*

Young et al - studies of ex-servicemen with unilateral brain injuries. Tested three things:

(a) Familiar face recognition

(b) Unfamiliar face matching

(c) Analysis of emotional face expressions

Found selective impairments in each of these abilities - e.g. some participants with right hemisphere damage were impaired for (a), others for (b) only. Left hemisphere damage associated with (c) impaired only. Response latency data also indicates support for expression processing deficit being selective, but processing for (a) and (b) not completely independent.

Distinction between face recognition and awareness of recognition is important. **Bauer** - skin conductance response experiment on LF showed recognition occurred even though LF was not aware of recognising a face - and example of **covert recognition**. Led to **Bauer** proposing two separate neural pathways at work - one for conscious, the other for non-conscious recognition.

Capgras delusion - (belief a person is an imposter or an object is a double; rare that both occur simultaneously) may be a 'mirror image' of prosopagnosia. **Ellis and Young** - suggests it results from damage to the dorsal route (covert recognition route) instead of the ventral route (overt recognition), damaged in prosopagnosia. Prediction is that Capgras sufferers would recognise familiar faces but fail to show an emotional response to them. Support from several studies - e.g. **Hirstein and Ramachandran** - overt recognition intact; covert recognition damaged.

Sergent and Poncet - able to demonstrate provoked overt recognition possible - a possible way of helping those with Capgras delusion. PV shown 8 faces of famous people; unable to recognise them. Told they all had the same occupation; PV was able to say they were politicians, name 7/8 and recall bio detail about the 8th.

IAC model - covert without overt recognition is the weakening in connections between FRUs and PINs. Face seen; FRU activated, but weakened FRU-PIN connection means threshold not crossed to recognise face overtly. Telling PV that the faces are related is the equivalent of strengthening PIN-SIU connections; activation is passed back to the PINs and the overt recognition threshold is passed. Provoked overt recognition successfully modelled in IAC by **Morrison et al**.

Are faces 'special'?

Questions:

1. Is there a region of the brain that underlies face processing - and if there is, does it mean face processing is qualitatively different from the processing of other visual stimuli.

2. Is face processing innate or learned?

3. How important are individual features, the relationship between them and the 3D structure of the face to recognition - and are they processed individually or as a whole?

1. Prosopagnosia can leave object recognition intact but damage face recognition and vice-versa - a double dissociation. fMRI scans show facial stimuli activate a particular area of the brain but it is not activated for other stimuli (beyond human/monkey faces.) Therefore, there is evidence to suggest there is specialisation of brain areas for face recognition.

2. Innate ability to process faces is supported by studies of newborns - but **Johnson and Morton** argue that there is a mechanism that makes newborns attend to faces rather than humans having an innate neural mechanism that processes faces. We have an innate attentional bias, which then serves to guide subsequent learning.

Putzar, 2010: Experimental studies of 15 Ps who had impaired vision through cataracts in early infancy show they can recognise faces as well as control subjects under normal lighting and orientation; but if faces presented at different angles and different lighting less perform less well than controls. => innate capability but supplemented by attentional mechanisms early in development.

3. Face recognition may be special as faces all tend to look alike if they have similar features in similar positions. Support comes from studies that examine the **inversion effect**. **Yin, Johnston et al** show that inverting a photograph of a face disrupts recognition more than inverting a photograph of an object. **Yin** - experiments show recognition memory for faces is better than objects when the right way up; opposite is true for inverted stimuli => faces processed differently

N170 ERP - face inversion delays it and increases activity in object regions … inverted faces treated as objects presumably because of the problem of detecting first order relations - **Maurer, Le Grand & Mondloch.**

Diamond and Carey - alternative hypothesis - our perceptual mechanism is tuned to see upright faces; tuning is lost when faces are inverted. Investigated dog experts vs non-experts. Shown photographs of dogs and faces; asked to memorise. All participants recognise upright faces better than inverted faces; however, dog experts recognise upright dogs better than inverted dogs. The inversion effect may therefore be acquired through expertise, so is not a face specific effect.

Their explanation as to what changes as we acquire expertise is a distinction between first-order (spatial relationships between the parts of the face) and second-order relationships (more subtle differences in the basic configuration, such as eyebrow shape/thickness etc.) Expertise => greater sensitivity to second-order properties. **Searcy and Bartlett** – grotesque faces presented upright and inverted. Faces rated less grotesque when shown inverted when spatial relations between features distorted; but no change in ratings if distortions were performed on the actual features themselves. Supports **Diamond and Carey** hypothesis. Also suggests (upright) faces processed as 'configurations' rather than collections of distinct features – **configural processing**. Thatcherised faces (**Thompson, in Maurer, Le Grand and Mondloch**) – only look grotesque when the right way up – don't appear so upside down. As rotated from upside-down to normal orientation, fMRI markers show changes in areas correlated with bizarreness/grotesqueness. On balance, although evidence exists for supporting the notion of specialise areas for face processing and that there is an innate ability to pay attention to faces, the processes of face recognition are probably not unique. **Conclusion** Still considerable research required before a comprehensive/detailed theory of recognition can be created. Difficult to do, as there are different types of recognition and different modalities may be used => these differences may involve different processes.		