Introduction

Errors that we make in perception - e.g. Müller-Lyer, Necker cube, Kanizsa's illusory square help us to understand the sophistication of the cognitive processes that permit visual perception.

Rationale - if they were simple, rather than sophisticated processes, then these illusions wouldn't occur, as they are all based on simple shapes. The likelihood is that all three illusions occur because we are attuned to deal with 3D objects and scenes - they occur because we try to interpret 2D shapes in this way.

Perceptual errors in cases of brain damage - e.g. prosopagnosia, neglect, Capgras syndrome etc. also demonstrate the complexity involved in visual perception.

Perceiving and sensing

Definitions of perception usually include the idea that it involves the analysis of sensory information. However, some philosophers reject that sensation plays any part in object perception – **Atherton (2002)** suggests that perhaps this is because they seem to be an unnecessary intrusion between a round dish and our perception of it as round.

Working definitions of: 'sensation' = detection of light, sound, heat etc. by our sense organs; 'perception' = the process of us creating a description of our world.

Much research has been done on visual perception; less on other modes - as we generally rely on vision most. Used to help us avoid and recognise objects as well as reading and face recognition.

The eye

The retina is made of rods (work with low light levels) and cones (responsible for colour and definition). We have a lower ratio of rods to cones than most animals.

Chapter3 - Perception

Approaches to perception

Different approaches taken - one is to consider if the 'goal' of perception is action or recognition. Not the same - sometimes we need to take action to stop something hitting us before we know what the object is.

Evidence exists that different neural mechanisms are involved in these aspects (**Milner & Goodale, 1998)** But they are not completely separate either.

A second way of differentiating approaches is on the 'flow of information' - 'bottom-up' - from the image formed on the retina until we have a representation of what we see, or 'top-down' - by starting with our existing knowledge of the environment and using that to guide the processing of sensory information.

Gestalt approach to perception

First part of C20. Philosophy: "the whole is greater than the sum of its parts". Therefore, an image is perceived in a particular way because of the organisation of the elements it is made of and not because of the nature of the elements themselves.

Perceptual organisation - a powerful phenomenon. Gestalt principles include **closure** (not quite closed circle perceived as closed); **good continuation** (two crossing lines perceived rather than two touching tips); **proximity** (a square made up of dots appearing as vertical or horizontal lines depending on the closeness of the dots in these orientations); **similarity** (a square made up of different colour dots appearing as vertical or horizontal lines depending on the organisation of the colours.) Two organising principles may conflict with each other and you might perceive either according to similarity, proximity and be able to move between the two perceptions.

Law of Pragnanz (**Koffka)**: "of several geometrically possible organisations that one will actually occur which possesses the best, simplest and most stable shape"

There are Gestalt demonstrations that do have equivalents "in real life" - e.g. a football occluded by a post - the whole sphere is perceived.

However, as 2D demonstrations are used, this approach to explaining perception is not without controversy. Others believe that visual perception can only be explained by studying how real 3D objects are perceived in the context of a complex 3D environment.

Gibson's theory of perception

A bottom-up approach - based on the premise that all information required for perception comes from the senses - no cognitive processing is required. Gibson argues that understanding perception is simply about understanding how (sensory) information is "picked up".

Strong link between perception and action - and action is the 'end point' of perception rather than just the creation of an internal description of the environment.

Perception is **direct** - information in light is sufficient to allow someone to interact with the environment. The consequence of this is that the perception of a 2D image is indirect - 'it's not a pipe, but an image of a pipe'.

An ecological approach

- as Gibson's theory places importance on perception in the 'real world' and it is that world that structures the light that reaches our retinas.

Driver for the theory came from training pilots to take off and land. Good 'depth of perception' is required. Tests based on 2D stimuli did not predict the performance of a pilot landing (**Gibson, 1947**).

To explain perception, it is necessary to understand and explain how the complicated surfaces and textures present in the real world provide information to us about our environment. **Gibson** - a "ground" theory (vs "air" theories, based on isolated artificial flat (plane) shapes): a surface is substantial; surfaces are textured; surfaces never perfectly transparent; surfaces can be seen (planes are the opposites of this.)

The optic array and invariant information

The structure imposed on light by surfaces around us was termed as the **ambient optic array** by **Gibson**. Consists of a series of angles of reflected light from surfaces. The optic array changes as you move your head around (side to side; up and down etc.)

Supplemented by higher-order features termed **invariants** - observers perceive information from the world by sampling the optic array to detect invariants. e.g. **Sedgwick, 1973** - 'horizon ratio relation' - how much an object is above the horizon to how much appears below remains constant (invariant) as the object gets closer or further away from the observer.

Another key invariant is the **texture gradient** - three key forms, relating to density, perspective and compression of texture elements. How it is created varies by surface type - e.g. in a carpet by the twists of material; on a road by the stones that make up the surface. We make an assumption that texture is uniform - i.e. a road consists of similar sized stones along its length. Therefore, changes in texture gradient provide information about distance, orientation and curvature of the surface we perceive.

If a surface is receding, the density of texture elements will increase with distance. The perspective gradient (width of elements) and compression gradient (height of elements) help us to detect shape and orientation of a surface. It is lack of texture information that causes ambiguity in what is perceived and therefore helps to explain phenomena like the Necker cube illusion.

Flow in the ambient optic array

Gibson - frozen structure does not (commonly) occur - invariants of structure do not exist except in relation to variants.

Another criticism of lab experiments is that motion is often absent - and he is arguing that invariant information can only be "picked up" in relation to a changing environment.

Two forms of motion - that of the observer and that of objects in the environment. The first usually produces the most movement - the entire optic array is transformed, which provides information about the positions and shapes of surfaces and objects.

Shape and position is revealed by:

(i) **motion parallax** - the further away something is, the less is appears to move.

(ii) **occlusion** - motion of the observer causes objects that are further away to be occluded by closer ones.

Motion of the observer causes **flow patterns** in the optic array. **Gibson, 1979**, proposed a set of 4 rules to link flow in the optic array to the movement of the observer.

(i) Flow in the ambient optic array implies the observer is moving: no flow = no movement of the observer.

(ii) Outflow of the optic array from the **pole** means the observer is moving towards it; inflow implies movement of the observer away from the pole.

(iii) Direction of the pole specifies the direction the observer is moving in

(iv) Change in direction of the pole => observer is moving in a new direction.

Therefore, movement is central to Gibson's interpretation of what perception is. Not limited to the eyes and other senses; but our *perceptual system* is a hierarchy - our head can turn, linked to a body that moves. He concludes that perceiving is an act (of attention) rather than a (triggered) response; an achievement, not a reflex.

Affordances and resonance

Gibson's theory goes beyond suggesting that perception = information "picked up" from the environment and argued the goal of perception was not a description of the world, but that objects directly afford their use.

Least controversial aspect is that it builds on the ideas of the Gestaltists - features of objects provide clues as to their use.

Two further claims **Gibson** makes about affordance are more controversial:

(i) Affordances are a bridge between perception and action, therefore cognitive processes are not required to intervene between the two.

(ii) There is no role for memory in perception. Instead, the perceptual system **resonates** to invariant information in the ambient optic array.

Other researchers counter (i) by arguing the cognitive processes that allow perception must be one focus of attention and (ii) if we don't use our memory to learn from previous mistakes, we would never learn from them.

Gibson's theory has been influential - and if nothing else, Wade and Bruce note that his criticisms made of artificial stimuli are still valid - we should not be blind "to the differences that exist between the virtual and the real."

Marr's theory of perception

Attempts to address how the brain "picks up" information from the environment - a criticism of **Gibson**'s theory is that this is not well explained.

Similarity with **Gibson** - information from the senses is sufficient for perception to occur

Main difference - an 'information processing' approach - processes responsible for analysis the images that reach the retina are central to perception.

Marr's theory is therefore strongly 'bottom-up': retinal image is the start point and then explores how this is analysed to produce a description of our environment. *Focuses on the perceptual processes used for object recognition*.

Work concentrates on computational theory and algorithmic levels of analysis - not biological/neural in focus - largely ignores the hardware of the brain.

4 stages of perception:

(i) **Grey level description** (light intensity at each point on the retina)

(ii) **Primal sketch**
- (a) *Raw primal sketch* - areas that are the edges & textures of objects identified
- (b) *Full primal sketch* - areas identified generate a description of the outline of objects.

(iii) **2½D sketch** (how surfaces relate to each other and the observer)

(iv) **3D object-centred description** (allow object to be identified from any angle)

(i) Grey level description

As the theory assumes perception is modular, the first stage of perception concentrates on the intensity of light at all points on the retina (colour is handled by a separate module later on.) - a greyscale representation. Produced by the pattern of depolarisation on the retina - caused by different potentials across the cell membrane.

(ii) Primal sketch

Two sub-stages used. First is to form the raw primal sketch by finding patterns in the grey level description. **Marr and Hildreth** (1980) using an algorithm implemented on a computer - **Gaussian blurring** - demonstrated how comparing images blurred to different degrees could enable the identification of object edges, based on blobs, edge-segments, terminations and bars.

There is evidence that retinal processing really does produce images have been blurred to different degrees.

The raw primal sketch is transformed into the full primal sketch - this contains information about the organisation of the image, including location, shape, texture and internal parts of objects that are visible.

Constructed from **place tokens** that are assigned to parts of the raw primal sketch based on how edge-segments etc. are grouped. Groups of place tokens are then aggregated to form higher-order place tokens.

This process is achieved by **clustering** - analogous to the Gestalt principle of proximity and **curvilinear aggregation** - analogous to good continuation. In other words, **Marr** saw the algorithms that turn a raw primal sketch into a full primal sketch as the implementation of laws such as those expressed by the Gestalt approach.

(iii) The 2½D sketch

To specify the layout of surfaces, cues that provide information on how far away each surface is need incorporating.

The modular nature of the process Marr describes means that other visual information is analysed at the same time the primal sketch is being created. For example, information associated with depth perception, motion, texture and shading cues.

Information from all these modules is then aggregated to form the 2½D sketch - the specification of position and depth of surfaces is relative to the observer - i.e. **viewer-centred**, containing no information that is not present on the retinal image. **Marr** describes it as a series of primitives with vectors that show the orientation of each surface.

Process of turning 2½D sketch to a 3D **object-centred** description is in (*Ch.4*).

Evaluation of Marr's approach

Research consistent with the mechanisms proposed

Marr and Hildreth (1980) tested the mechanism that creates the raw primal sketch through Gaussian blurring by using a computer program - results demonstrated that it was successful at finding the edges of objects [but it doesn't necessarily follow that this is the mechanism used by our perceptual system.]

Marr's theory fits experimental results when considering the integration of depth cues in the 2½D sketch. **Young et al** (**1993**) used an experiment that isolated motion and texture cues and concluded that the perceptual system does process these separately and selective use is made dependent on how good (less 'noisy') they are.

Research that is not consistent with Marr

Enns and Rensick (1990) - showed people can easily determine the odd one out in a series of block figures where the only difference between blocks is orientation. Implies the mechanism proposed by **Marr** for the creation of the full primal sketch is faulty - as grouping strategies must make use of 3D information.

The theory can be evaluated using broader concepts too - for example, the neurophysiological evidence that there are two visual pathways in the brain - one for 'where' information and the other for 'what' information. There also appear to be different perceptual processes depending on our perceptual goal - action or recognition.

Marr's theory predicts neither of these aspects - **Wade and Bruce** (**2001**) argue that the separation of visual pathways into one for action and one for object recognition is very difficult to incorporate into the theory.

The impact of the theory has been significant - **Wade and Bruce** suggest *it is not the details - but the (modular) approach itself that is the most valuable aspect* of **Marr**'s theory.

Constructivist approaches to perception

Notion that the sensory information that forms the basis of perception is incomplete - so we therefore have to construct our perception of the world from what we already know and what is sensed. *Stored knowledge is therefore used to help us recognise objects.*

This approach is most closely associated with **Rock** and **Gregory**. **Gregory** argued we try to recognise things by generating a series of **perceptual hypotheses**. This is required as sensory data are not complete (and so if we did have complete, perfect sensory data we wouldn't need to generate such hypotheses.)

Stored knowledge is therefore used to enable the construction of these hypotheses. The usefulness of such stored knowledge has been demonstrated by using impoverished figures - e.g. **Street, 1931** - an outline of a liner that is difficult to perceive until you are told what it is. Sensory data hasn't changed but knowledge has to enable this perception to occur.

Having knowledge doesn't always lead to a correct perception. Some things (for example, faces) are so familiar that we are biased towards one particular hypothesis. Can be demonstrated by looking at a hollow face, such as an Egyptian burial mask. This runs contrary to our expectations, so much so that even our high-level knowledge that the mask is hollow still doesn't stop our perception of it as being a normal face. **Gregory** suggests it represents our tendency to accept the 'most likely' hypothesis.

The 'impossible' **Penrose** triangle illustrates a similar point - attention keeps being drawn to each corner in turn so that we perceive it as being a credible 3D shape. This apparently data supported hypothesis drowns out the knowledge we are looking at a flat pattern.

Areas of **Gregory**'s theory are vague:

(i) How do we know when we've reached the right hypothesis and so stop generating new ones?

(ii) Why does knowledge help perception - but not always?

(iii) How can we know something is wrong and yet still perceive it incorrectly - e.g. in the case of a hollow face?

However, there is evidence that our perceptions are 'constructed' from bottom-up and top-down information.

Gregory's theory appears to be in direct conflict with **Gibson and Marr**'s theories - bottom-up only. There is however the *potential to reconcile the three when considering the way in which the brain actually appears to process information*.

The physiology of the human visual system

Shapley - there are at least two semi-distinct streams of information from the retina, via the optic nerve, into the brain. They do however overlap in the type of information they carry and have many connections between them - but they can be conceptually thought of as being distinct.

From the eye to the brain

Two distinct streams - the parvocellular and magnocellular pathways - are apparent even at the level of retinal ganglion cells. This is so into and within the primary visual cortex (V1), but there are interconnections between the two.

From V1 onwards information is still in two streams - the **ventral stream** to the inferotemporal cortex and the **dorsal stream** to the parietal cortex.

Dorsal and ventral streams

Ventral stream - projects to brain regions involved in pattern discrimination & object recognition.

Dorsal stream - projects to brain regions that deal with the position and movement of objects.

Schneider (**1967,69**) - work on hamsters - suggested two distinct parts of the visual system - one part for pattern discrimination, the other for orientation in space. The 'what is it' and 'where is it' systems. Later research (**Ungerleider & Mishkin, 1982**) led to the ventral stream being termed a 'what' system with the dorsal stream termed a 'where' system.

However, there is lots of evidence that the streams are interconnected and that they appear to converge at the pre-frontal cortex (**Rao et al**.), but with some evidence their distinction is maintained (**Courtney et al**.)

Milner and Goodale report on DF - apparently unable to use her ventral system for analysing sensory input. Not able to recognise faces or objects or see the difference between simple geometric shapes. DF could draw objects from memory but then not recognise them. Her dorsal stream did appear to be intact as she was able to pick discs up of various sizes, with the distance between her finger and thumb well correlated with the size of the disc - even though it wasn't possible for her to say if two discs were the same or of different sizes.

Therefore, size information was not available to DF's conscious perception (ventral stream) but it could guide action (dorsal stream.)

Norman (**2002**) proposes a dual process approach. The streams act synergistically - but dorsal stream = perception for action, ventral stream = perception for recognition. Support from:

1. **Goodale and Milner**, **Ungerleider and Mishkin** - ventral = recognition; dorsal = visually guided behaviour (pointing, grasping …)

2. Ventral = processing of fine detail (**Baizer et al**); Dorsal = processing of motion (**Logothesis**)

3. DF studies - ventral = knowledge-based; dorsal = very short term storage only (**Bridgeman et al.**)

4. Dorsal receives information faster than ventral (**Bullier and Nowak**)

5. Some psychophysical evidence => we are more conscious of ventral than dorsal stream functioning (**Ho**.)

6. **Goodale and Milner** - ventral stream is object-centred (as it recognises things) and dorsal stream drives action - i.e. viewer centred (*See Ch. 4*)

Rel'ship between pathways and theories of pceptn.

Gibson - perception for action (Dorsal stream)
Marr - perception for recognition (Ventral stream). Constructivist approach is also mainly concerned with perception for recognition.

Gibson's notion of "affordance" is concerned with what things are *for*, not what they really *are*. In other words, linked to actions - and therefore potentially to the dorsal stream.

Gibson's ecological approach (no need for memory in perception) is again similar to what we know about the dorsal system - **Bridgeman et al**, **Creem and Proffitt** - found only a very limited amount of memory associated with the dorsal system.

So, evidence to suggest the dorsal stream is Gibsonian in operation.

Ventral stream appears to help with object recognition - e.g. processing the fine detail **Marr** believes is required to tell different objects apart. Also appears to be able to draw on top-down knowledge to aid identification.

Having both makes sense - if you need to be avoid being hit by something, you just have to move to avoid it - you don't really need to identify it first!

A dual-process approach?

Possibly a danger of trying to make what is known about the dorsal and ventral streams fit into the framework of **Gibson**/**Marr**/Constructivist theories.

Vagueness in both Gibson and Gregory's theories on how the processes they postulate are implemented makes it questionable about how good a framework they are to interpret the workings of the streams.

These theories also emphasis the difference between the streams - yet it is known they are highly synergistic and interconnected. (**Norman, 2002**)

Interaction between dorsal and ventral streams (**Binstead and Carlton**) - an illustration is the skill of driving. Initially, cognitive processes of the kind associated with the ventral stream are required, then as the skill is acquired, it is more similar to capabilities associated with the dorsal stream.

But how would such a transfer occur - and is the case instead that the two are so closely linked they should not be considered at all separate in the first place?

Processing may therefore be not an either/or proposition, but it could be for action and recognition.

Combining bottom-up and top-down processing

Perception is likely to contain elements of both types of processing.

Standard explanations of backward masking need the mask to contain overlapping contours or exactly coincide with those of the target (**Enns and Di Lollo**.)

However, **Enns and Di Lollo** demonstrated a four dot pattern that did not intersect the contours of the target also masks. This was explained by citing **re-entrant processing**. Neuroscience shows if one brain region is sending a signal to another, then it sends a signal back through re-entrant pathways (**Felleman and Van Essen**.) **Hupe et al** suggest these are used to allow the brain to check a perceptual hypothesis. But if the information coming from vision changes too rapidly, **Di Lollo** argues that as the target will have been replaced by the mask, the perceptual hypothesis being tested will be rejected by the new bottom-up information.

This explanation is therefore based on the premises that top-down and bottom-up processes interact. Consistent with the idea that perception needs both top-down and bottom-up information.

Conclusions

(i) Perception is complex - even for simple objects.

(ii) Many influential theories exist.