Visual perception and attention

1. Visual perception

There is no doubt that seeing is an amazing part of the human experience. Putting aside the mechanics of the eye, what is it we 'actually see' is not exactly easy to understand. Our ability to create mental imagery – to “see with the mind’s eye” – has been of interest to philosophers and scientists for centuries.

History

The earliest attempts towards studying visual theories date back to 400-300 BC where Plato proposed that “the soul is the source of vision, with light rays emanating from the eyes and illuminating objects” (Lindberg 1981). Later, in 999 AD, Alhazen, also known as Abu Ali Hasan Ibn al-Haitham (present-day Iraq), used spherical and parabolic mirrors to study spherical aberration and gives the first accurate account of vision--that the eyes receive light, rather than transmit it. Alhazen also investigated magnification resulting from atmospheric refraction and writes about the anatomy of the human eye and describes how the lens forms an image on the retina in his famous major optical work “Opticae Thesaurus” (Opticae thesaurus Alhazeni libri vii), the first real contribution to the science of optics in the first millennium (Lindberg 1981).

In the modern times, a German polymath, Hermann von Helmholtz introduced the notion that visual perceptions are unconscious inferences. For him human perception is but indirectly related to objects, being inferred from fragmentary and often hardly relevant data signalled by the eyes, so requiring inferences from knowledge of the world to make sense of the sensory signals (von Helmholtz 1866).

On-going research

In the evolution of theories of visual perception psychologists consider many different approaches, from Gestalt theory and Brunswick probabilistic functionalism to Gibson’s direct perception and ecological optic and Marr’s computational approach. However, Ian Gordon (Gordon 224-229) reasons that there is a noticeable lack of agreement between theorists over areas such as:

- The definition of Stimulus
- The appropriate level of explanation
- The place of subjective experience in perceptual theory
- The evolutionary background to human vision

Although rivalry between different theories will continue, psychologist Richard Gregory states that ‘from the patterns of stimulation in the retina we perceive the world of objects (1977). How we explain the process that allows these “patterns of stimuli” to become recognizable objects is problematic.

Vision dominates our perceptual systems and is estimated to use 50 per cent of the brains cortex. Light enters the eye through the pupil, striking the receptors in the retina, causing them to respond by emitting an electrical impulse. The brain processes these electrical signals and forms representation (Visual Perception) of colour, shape, movement etc. (Palmer, 2002) (See Research Methods for HCI, Cairns, Cox et. al) The visual system is remarkable and capable of perceiving objects in sunlight and darkness, or rapidly moving, like darting insects; although cannot see a bullet, a plant growing or infra-red. (Preece, et al. 1995)

Gibson versus Gregory
There are two theories that aim to explain how the brain perceives the electrical signals it receives from visual stimuli in the eyes. Gibson (1966) takes an 'ecological' stance with the 'view' that information is simply detected, and we explore this through a combination of the senses. There is enough information around us to make sense of the world directly like, texture, objects brightness, movement, etc. The concept of 'affordances' (Norman, DOET) is an important aspect of this. The other theory is called 'constructionist' and suggests that what we see is a response to this stimuli 'the visual system constructs a model of the world by transforming, enhancing, distorting and discarding information'. (Gregory 1970) 'The effect of this is to provide us with a more constant view of the world, so if we are walking down a street, buildings are seen as not moving and people as approximately the same size and shape despite the images we see'. (Preece, et al. Human-Computer Interaction)

Both of these theories suggest that the art of 'seeing' is more than pictures being presented in our minds, although both infer that we are active perceivers. The difference between these two perspectives can be summarised in 'top down' and 'bottom up' processing. Many of the experiments that seem to support the top down theory, are done 'out of context' and it appears that both these ideas may go some way to explain how we 'see'.

**Simple Psychology website**

There is a good website that discusses these two theories some detail with interesting examples and pictures. (See [http://www.simplypsychology.org/perception-theories.html](http://www.simplypsychology.org/perception-theories.html))

'**Bottom-up processing**' is also known as data-driven processing, because perception begins with the stimulus itself. Processing is carried out in one direction from the retina to the visual cortex, with each successive stage in the visual pathway carrying out ever more complex analysis of the input.

There are some interesting examples of how Gibson argues against the top down theory and his arguments for surrounding context assisting with how we interpret information around us for example, 'Optic Flow Patterns' allow us to know if we are moving towards or away from and object, or the role of 'invariants' i.e. the flow of texture always occurs the same way around. 'Affordances' are cues in the environment that aid perception. In DOET, Norman describes affordances as the qualities of objects or environments that help users perceive their function and perform actions on them.

'**Top-down processing**' refers to the use of contextual information in pattern recognition. For example, understanding difficult handwriting is easier when reading complete sentences than when reading single and isolated words. This is because the meaning of the surrounding words provide a context to aid understanding.

There is a good example of how top down processing can affect visual perception on this site. Staring at the Necker cube below the orientation can suddenly change or flip. This is without any change of visual stimuli. There is also a good example with a rotating Charlie Chaplin mask that when you see it from behind, your tend to perceive the convex nose as concaved, supporting the need for the mind to construct the face as we are used to seeing it. Also see the example of a picture that once you look at it and read the text then takes on a very strong image, suggesting that we learn to see the image in a particular way.

**2. Biology of the visual system**

Being able to perceive the world is the result of light passing into the eye and being projected onto the retina. Several systems regulate how an image is presented to the retina. Firstly, light passes through a transparent layer of the eye called the cornea, which acts as a protective layer. Then, the amount of light allowed into the eye is regulated by the iris that adjusts the size of the pupil. Finally, the light passes through the lens which focuses the image (See figure 1). The retina contains millions of receptor cells known as photoreceptors, of which there are two kinds: cones and rods. These

Cones are the receptor cells that are sensitive to colour vision and rods are most sensitive low intensity light, therefore they help us to see at night or in low light conditions. When a change in the level of light is detected by the photoreceptor, it sends a message in the form of a neurotransmitter to the bipolar cell, which activates the ganglion cell which is responsible for sending the message to the brain through the optic nerve. The optic nerves from each eye meet in the brain at a section known as the optic chiasm. Here the information from both eyes converges and is split up into information for the left and right visual field (As shown in figure 2). Subsequent processing of information from the left visual field is done in the right hemisphere and information from the right visual field is done in the left hemisphere (Gazzaniga, 2002).

The optic nerve passes the information from the photoreceptors to the lateral geniculate nucleus (there is one of these in each hemisphere), which has six layers for analysing information from different areas of the visual field. Layers 2, 3 and 5 receive projections from the ipsilateral eye (on the same side of the body) and layers 1, 4 and 6 from the contralateral eye (on the opposite side of the body). Each layer receives information from different types of ganglion cells in the retina. The lateral geniculate nucleus organises and then outputs this information to the primary visual cortex (V1) via the optic radiations. Further processing of the retinal image is done here. V1 is the first processing area of the striate cortex. There are several layers that are responsible for processing different types of visual data, for example one area would be sensitive to variations in colour and another area for movement variation. As the information passes through the visual system, the data is integrated and begins to form recognisable objects.

The visual data continues in one of two directions (as shown in figure 3). The ventral stream passes information to the inferior temporal cortex which determines what an object is. Alternatively, information is sent to the visual association cortex via the dorsal stream which determines where an object is (Carlson, 2004).

Because of the complexity of the visual system, there have been some interesting case studies recorded. One case study worth noting is the case of the colour blind painter. This was recorded by Sacks (1995). In this case, an artist lost the ability to see any colour other than shades of grey after a car crash. Colourblindness is most usually something someone is born with and it is extremely rare that a person cannot see colours at all. A defect with a person’s cones is usually the reason for colourblindness. This colourblindness caused by brain damage, or cerebral achromatopsia, affected the artist's life significantly. Food lost it's appeal, he avoided social interaction and even lost interest in sexual intercourse. One of the shocking things about this case was, that even when he closed his eyes to eat, the mental image of his food would be as grey or black as it appeared to him. It became apparent that he could not even imagine or dream in colour. This was due to the loss of the ability to use V4. This area is responsible for higher order colour generation. Although he could still use the his wavelength sensitive V1 area which meant he could still see variation in lightness and darkness. This case study is documented in Sacks' An Anthropologist on Mars (1995). It's very interesting!

The Island of the Colour-blind (1996) is another book by Oliver Sacks which he describes his visit to the island of Pingelap in Micronesia. On this island, 5-10% of the inhabitants have a hereditary condition which means their cones are not functional.
3. Colour perception

Perception is the means of transcoding environmental information, through our different "input" devices, i.e. eyes, mouth, fingers, ears, into experiences of objects, sounds, events, tastes. [Roth:Rodgers]. Colour perception refers to the way our eye and our brain work together, in order to distinguish the full visible spectrum in order to identify colours and subsequently, the objects. A viewer is able to see an object when some of the light that hits it is reflected back. So, for instance, we perceive that many plants are green because of the pigment Chlorophyll (that provides energy to the plant for photosynthesis) absorbing red, yellow, blue, and violet wave lengths but reflecting green. Seeing colours is a complex process, and involves our eyes, our brain, the light, and some physics. For the physicist colour is a wavelength of light that an object either generates or reflects. This means that we use the language of the physicist to describe sensory stimuli we perceive as colour. For the psychologist, seemingly a colour like red suggests an internal process that may or may not be associated with an external event. If you close your eyes and think of a nice big juicy tomato, you may be able to see red even though this isn't down to an external stimulus, and proves that colour is also within us. We may consider therefore that colour isn't just dependent on the external world, but also originates through the power of imagination of our inner world. (Some of this is taken from Frank H. Mahnke, Colour, Environment and Response)

Let's Talk Nanometers, Wavelength and Frequency.

Nanometer is a unit of measurement, equivalent to one billionth of a meter. It is widely used for measurement in an atomic scale. In this case, we use it to specify the wavelength in which every single colour is visible to the human eye.

Definition of Colour

Simply put, colour is light of a variety of wavelengths and frequencies [colourtherapyhealing.com]. Isaac Newton discovered that colour colour was present in white light. He experimented with prisms to separate incoming white light into the colour spectrum and back to white light. (Wikipedia) This is famously used as the artwork for Pink Floyd's album "The Dark Side of the Moon". In this cover, a prism is used to "divide" the colours from the white light. This separation, reveals the six basic visible colours, i.e. Red, Orange, Yellow, Green, Blue, Violet. This process is called "refraction", and is the result of the change of speed caused to the white light, when it changes medium. [colourtherapyhealing.com].

The Visible Colour Spectrum:

<table>
<thead>
<tr>
<th>Colour</th>
<th>Frequency</th>
<th>Wavelength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Violet</td>
<td>668–789 THz</td>
<td>380–450 nm</td>
</tr>
<tr>
<td>Blue</td>
<td>631–668 THz</td>
<td>450–475 nm</td>
</tr>
<tr>
<td>Cyan</td>
<td>606–630 THz</td>
<td>476–495 nm</td>
</tr>
<tr>
<td>Green</td>
<td>526–606 THz</td>
<td>495–570 nm</td>
</tr>
<tr>
<td>Yellow</td>
<td>508–526 THz</td>
<td>570–590 nm</td>
</tr>
<tr>
<td>Orange</td>
<td>484–508 THz</td>
<td>590–620 nm</td>
</tr>
<tr>
<td>Red</td>
<td>400–484 THz</td>
<td>620–750 nm</td>
</tr>
</tbody>
</table>

The eye is not able to see the full spectrum, due to its physiological limitations. For instance, we can't distinguish the infrared and ultraviolet wavelengths. An example of an infrared wavelength is that little lightbulb in front of the remote. Every time we press a button it emits infrared light, invisible to the human eye, containing relevant information. The TV set has a receiver able to decode this information. The centre of gaze in the eye is called the fovea and in this area the cones are very tightly packed and provides the highest visual acuity. The area around the fovea has a mix of cones and rods and is used for gaining a broader visual scene and being aware of bigger objects.

Trichromatic theory

Thomas Young and Hermann von Helmholtz proposed the trichromatic theory in the 19th century and this describes how the eye can detect colours. The cones provide the basis for our colour perception and in general terms can be described as providing trichromatic colour vision, having different peaks of spectral sensitivity roughly corresponding to short, medium and long wavelengths of light. This doesn’t entirely correspond to blue, green and red so often used to conveniently label the cones. Colour perception is achieved by a complex process that starts with different outputs of these cells in the retina and finalised in the visual cortex and other areas of the brain. In order for us to see, an image has to be presented on the retina to stimulate these photoreceptors to send impulses up to the visual cortex. This is a constant stream of different coloured light photons entering the eye, reaching the retina and all the photo receptors firing their responses as they are stimulated, giving us an on going ever changing picture.

<table>
<thead>
<tr>
<th>Cone type</th>
<th>Name</th>
<th>Range</th>
<th>Peak wavelength</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>?</td>
<td>400–500 nm</td>
<td>420–440 nm</td>
</tr>
<tr>
<td>M</td>
<td>?</td>
<td>450–630 nm</td>
<td>534–555 nm</td>
</tr>
<tr>
<td>L</td>
<td>?</td>
<td>500–700 nm</td>
<td>564–580 nm</td>
</tr>
</tbody>
</table>

Additive colour theory

Red Green and Blue are the primary colours of white light and white is produced when these colours are present in equal amounts. James Maxwell first described the additive colour theory in the mid 1800s. Different combinations of these colours produce more colours. Blue and green produces cyan, blue and red produces magenta, red and green yellow. These are additive secondary colours.
The beginning of colour perception

As light strikes the retina, light energy is converted into neural activity. The retina is several thin layers of complex cells where connections amongst the various neurons are made. The retina is often thought of as an extension of the brain as it has a processing capability. The deepest layer of these neurons are the photoreceptors, located at the back of these layers and these are the only cells in the retina that can translate light into nerve impulses. Amazingly the light has to pass through all the other layers before it reaches the photoreceptors. As the layers are very thin and transparent they don't blur the image that much and this doesn't really matter for our peripheral vision. However the cones are exposed in the fovea, with the other cells displaced around in a ring to ensure a sharp image falls on the cones.

The process uses light sensitive pigments located on disks around the rods and cones. When light hits these pigments they change causing a cascade of chemical reactions and nerve signals are then sent to the next layer of cells in the retina, the bipolar-cells and then the ganglion cells. The ganglion cells have axons or nerve fibres that form the optic nerve and carry the impulses towards the brain. There are other cells in the retina that contribute to the processing of visual information and receive or transmit impulses.

‘The gigantic job of taking the light that falls on the two retinas and translating it into a meaningful visual scene is often curiously ignored, as though all we needed in order to see was an image of the external world perfectly focused on the retina. Although obtaining focused images is no mean task, it is modest compared with the work of the nervous system—the retina plus the brain’

Opponent colour process

Ewald Hering proposed opponent colour theory in 1892. This theory defined that the human visual system interprets information about colour by processing signals from cones and rods in an antagonistic manner. That is, it records the differences rather than individual cone responses and it is proposed that this is more efficient for processing. The three opponent channels are red verses green; blue verses yellow; and black verses white for light and dark or luminance. The opponent process is thought to be at a higher processing level, taking the signals from the three types of cone and turning them into colour perception as a basis for seeing visual images. The bipolar and ganglion cells are thought to be related to the opponent process. One visual characteristic that is often sited to support this theory is staring at a red square for some time and then looking away onto a white piece of paper and then seeing a green square.

4. Perceptual organisation

Perceptual organisation theories aim to explain how human brain uses the information of the world, acquired through senses, to build the internal perception of the world.

The assumption is that senses provide a partial, ambiguous set of information: from the mechanical point of view, for example, the human eye is capable to see a restricted portion of the electromagnetic spectrum, and we can hear an interval of sound frequencies. Shape and position of our sensory organs provide additional constrains.

Despite our limitations we are capable to elaborate, catalogue, associate and use the information collected to interact with the environment that surrounds us.

Few general principles can be identified that describe how human brain deals with stimuli:

- coupling
- multistability
- parts grouped in a whole
- simplicity
- likelihood

The coupling principle states that when our sensors receive a stimulus, the stimulus is compared and associated to a matching element in our memory. Early radiophonic special effects were based on this principle, for example the sound of a thunder was generated oscillating abruptly a metal sheet. The sound heard was easily associated by the brain to a thunder since the closest match stored in the memory.

Multistability is present when the same stimulus can be associated to one or more interpretation. The Necker cube is an example of visual multistability, where the cube visual representation can be seen as bi-dimensional or axonometric. Another example of multistability is the rotating dancer, a rotating silhouette of a dancer, that can appear as rotating clockwise or counter-clockwise.

Parts often are grouped to form a whole that can be recognised more easily, and is often memorised as a whole (a face versus the nose/ eyes/ mouth). The grouping and its perceptual effects has been extensively explored by the Gestalt. The grouping of parts can also be reconduced to the recognition patterns where the structure is the matching criteria, not the single component.

The simplicity principle, developed by the Gestalt in the Pragnanz, states that when presented with a stimulus, the neural system tends to the lowest energy demanding solution, the simplest, most stable status or solution. According to Information Theory, the system tends to the solution requiring the least amount of information. Smallman and Cook demonstrated how realistic displays fail to provide efficient representation.

According to the likelihood principle, enunciated for the first time in 1909 by von Helmholtz, in front of a visual stimulus we tend to assimilate it to the object or situation that most likely matches the pattern of the confronted stimulus.

As can be evinced from the Likelihood and Simplicity principles, there is a tension between the two and a potential contradiction that has been extensively explored in literature and is often referred to as the contraposition of Occam's razor versus von Helmholtz.

Most principles apply to all senses, however their validity has been extensively explored for the visual stimuli.

Gestalt psychologists were among the first to analyse the visual perception and how relates to organisation of the knowledge, developing the simplicity principle.
The main attempt to study perceptual segregation and organisation was made by the Gestaltists. They worked off of one main principle, which states that “of several geometrically possible organisations that one will actually occur which possesses the best, simplest and most stable shape” (Eysenck & Keane, 2010, pp.70). This is known as the law of Pragnanz.

Four other laws of perceptual organisation were developed from the law of Pragnanz. The first law, the law of proximity explains how visual elements are usually grouped together if they are close to each other (Eysenck & Keane, 2010). So we can see in figure 1 that the dots on the left are perceived as one group while the dots on the right are perceived as three groups due to the law of proximity.

The law of similarity is the second Gestalt law. This law states that objects will be organised together if they look similar (Eysenck & Keane, 2010). So we can see in figure 2 that horizontal rows are perceived rather than vertical rows or one group of dots. This is due to the law of similarity.

The law of good continuation is observable in figure 3 as fewest interruptions are perceived by grouping this as two overlapping lines. Fewer people would perceive this image as four separate lines touching in the middle or two V shapes touching at their points as this would interrupt the straightness or curvedness of the lines (Eysenck & Keane, 2010).

The fourth Gestalt law is the law of closure. This states that incomplete shapes are perceived as just this, rather than separate lines (Eysenck & Keane, 2010). So in figure 4 the lines on the left are perceived as an incomplete circle, rather than four curves and likewise with the incomplete square. Likewise, the panda (figure 5) is perceived to be whole, even though there is no line indicating this.

Most of the Gestalt laws are focused on static two-dimensional shapes. There is one interesting law which states that visual elements that move together are grouped together (Eysenck & Keane, 2010). This is the law of common fate and can be observed by lights in a dark background. These lights seem to be separate unless they move together. This can be observed in this moving image: http://www.infovis-wiki.net/images/f/f9/Common.gif.

5. Object recognition

Visual Attention

General question: Is visual attention location or object based? Do we pay visual attention to the regions of space within a visual scene or directly to specific objects regardless their spatial context?

What we know for sure, visual attention is very selective. We have a finite amount of attentional resources – we couldn’t attend to everything if we wanted to!

Sternberg (1999) “Attention acts as a means of focusing limited mental resources on the information and cognitive processes that are most salient at a given moment”

Proposed analogies for location based visual attention - The idea is the same for all of them: objects that fall under the “beam” of attention are subject to further processing with priority.

- Spotlight (Posner, 1980) - Using spatial cueing tasks he concluded that moving attention is a cognitive phenomenon not tied to physical eye movements but instead, an internal mechanism. Attention can be likened to a spotlight that enhances the efficiency of detection of events within its beam.
- Zoom-lens (Eriksen & St. James, 1986) - We have ability to increase and decrease diameter of such spotlight beam.
- Multiple spotlights (Awh & Pashler, 2000) - Our attention can keep several spotlights in the visual field.

Evidence for object based visual attention - Attention selects from objects themselves, rather than potentially empty regions of space.

- Overlapping objects experiments (Duncan, 1984) suggests that objects, or groups of objects are parsed in accordance with Gestalt laws and are then subjected to further processing.
- Cognitive neuroscience studies on fMRI (Downing & Kanwisher, 2000; O’Craven et al, 1999) gives compelling evidence that attention can select individual objects, rather than whatever falls into a particular region of space.

General consensus on location vs. object based - Visual attention can operate in both ways, depending on the goals of attention.

Feature binding

Pop-out effect happens when one perceptual feature of the target is different (shape, colour, size,...). If features differs in conjunction, we need attention to glue these features and to apply serial search.

Feature Integration Theory (Treisman & Gelade, 1980)

- visual attention is location based
- we have feature maps for each type of visual feature + master map of locations
- feature detection occurs pre-attentively
- correct binding (search) requires serially applied focused attention

Proposed evidence - illusory conjunctions, Dual route - dorsal (spatial) vs. ventral (object) channels in brain, processing different visual features separately.

Object Files

Theory proposes that we're able to bind visual features into so called Object Files in order to maintain changes during relative movement without repeated binding:
Object representations (result of FIT - see above) are maintained over time and despite movement
objects are perceivable even while unidentified
we’re able of multiple object tracking (up to 3-5)
spatio-temporal continuity is critical more than content (changing the colour has no impact, changing the pace and direction has)

In order for objects to be tracked, they must maintain a spatio-temporally plausible path of motion, even if occlusion occurs. Object-files are sticky, but motion must be consistent with reality and expectations.

Object recognition

Models of object-recognition need to be able to allow accurate performance regardless of viewing conditions.

Viewpoint-Invariant Theories

These suggest that object recognition is based on structural information, such as individual parts, allowing for recognition to take place regardless of the object’s viewpoint.

- 3 stages representation model (Marr & Nishihara, 1983) - Evidence proposed by evolution or neuroscience (e.g. Visual Agnosia - seeing without recognition, The Man Who Mistook his Wife for his Hat)
  1. 2D (primal sketch) - raw shape recognition through the contrast of light
  2. 2.5D (viewpoint dependent) - detecting distance (through stereopsis), orientation in space, texture,... but representation of the object still depends on viewpoint at this stage
  3. 3D (viewpoint independent) - full 3D representation of the object obtained from memory, independent of viewpoint, can be mentally rotated

- Recognition by components (Biederman, 1987) - recognition occurs by matching a set of 3D primitives (geons) to a stored representation in visual memory and we only need first and second stage as opposed to Marr. While the model can account for much real objects, it could not account for performance with novel objects.

Viewpoint-Dependent Theories

Viewpoint-dependent theories suggest that object recognition is affected by the viewpoint at which it is seen, implying that objects seen in novel viewpoints reduce the accuracy and speed of object identification.

- Extrapolation from Canonical viewpoints - typical, representative images of objects stored in memory while experiencing them.

We must also account the context influence - e.g. an object in unusual angle might be recognised slower in contextually inappropriate background.

And research also strongly evidence specialised areas of the brain, independently processing specific objects - Fusiform Face Area (FFA), Parahippocampal Place Area (PPA),...

6. The Gestalt view

Introduction

Gestalt theory was introduced in the field of psychology in 1912. Its major contributors were Wesheimer, Koffka, and Kohler (Ware, 2003). Gestalt aimed to explain the way people perceive structures in the environment by examining psychological phenomena. Most theories of the time like structuralism and behaviourism were analysing an experience as a sum of components, which were studied in isolation. The proposed solution was encapsulating the solutions of the isolated elements (Wertheimer, 1938a). In contrast to such approaches, Gestalt theory does not face an experience as ‘numbers’ or sum of individual elements that can be combined, but as parts of the larger units, namely the ‘wholes’, separated from and related to one another (Wertheimer, 1938b). As Wertheimer (1938a) explicitly states:

"There are wholes, the behaviour of which is not determined by that of their individual elements, but where the part-processes are themselves determined by the intrinsic nature of the whole. It is the hope of Gestalt theory to determine the nature of such wholes."

As it is concluded, Gestalt theory adapts the notion that people tend to organise elements in large units. It can be considered a bottom-up approach in a sense that reaches complex cognitive processes starting from simple elements that stimulate the perception (Carlson and Heth, 2010). From another point of view, it can also be assessed as a top-down approach as it highlights the effect of the context and holism (Soegaard, 2010), which also hold a key role in the contemporary research in the field of Human-Computer Interaction.

Gestalt theory in Art and Visual Communication Education

What made gestalt theory appealing to visual artists, educators, and visual communicators is that the school of psychology sought to explain “pattern seeking” in human behaviour. In 1944 the New Bahaus in Chicago, used illustrations by Wertheimer, Koffka, and Kohler to assist his discussion of the laws of visual organization and psychological forces. Some major contributors have been Professor Rudolf Arnheim and Roy Behrens, who in 2002 incorporated Gestalt in her interactive media Design curriculum. (Graham, 2008)

Interactive Media Design and Gestalt
Even though there has been considerable literature on Graphic and visual design and Gestalt, little has been mentioned in connection to website and particularly to interaction design and its relation to its main principles. We hereby will venture to include Gestalt's correlation to interactivity and later on "Interaction Gestalt" as coined by (Lim et al, 2007) and its importance in HCI, aesthetics and User Experience.

Interaction Gestalt

"The crucial point, however, is to understand that the way we design an interaction inevitably impacts experience. In that sense, Yong-kyung Lim and colleagues (2007) offer the concept of interaction gestalt as a bridge between the interactive product and experience. They provide eleven attributes to describe those gestalts, such as pace, speed, proximity and so forth. The authors emphasise that these properties 'are not experience qualities—they are simply descriptions of the shape of the interaction' (Lim et al., 2007, p. 249)."

(Hassenzahl, 2010)

"Interaction Gestalt" has been coined by (Lim et al, 2007) in an effort to raise salience of aesthetics concept within the HCI science. Through Gestalt principles they propose a triad of questions to design gestalt interactions:

1. What is being designed?
2. What can be manipulated, of that which is designed? (i.e attributes)
3. How can these attributes be manipulated?

It is important to understand that the attributes are not supposed to be used individually. As the original meaning of gestalt tells us, the sum is different from the whole.

Designers should have knowledge of how to shape aesthetic interactions in a more visible, explicit, and designerly way. This is a kind of knowledge we are currently missing in HCI." (Lim et al, 2007)

Shaping the gestalt involves both imaging how the gestalt should be manifested in an interactive artefact as well as anticipating how users will experience the gestalt.

The interaction gestalt also has to be designed in a way that will evoke the desired user experiences....Traditionally in HCI, interactions have been described by languages of

1. interface styles such as WIMP (widows, icons, menus, and pointing device),
2. forms of interface devices such as tangible interfaces and graphic user interfaces (GUIs),
3. actions that are supported by interfaces such as instructing, conversing, navigating, and browsing [35], and
4. object-based concepts such as spreadsheet applications de- signed following traditional ledger sheet forms [35].

Although all these approaches have helped conceptualising and shaping interface designs, they have not directly supported the aspect of aesthetics when designing interactive artefacts. (Lim et al, 2007)

“The most significant benefit of introducing this concept for aesthetics of interaction is that it enables designers to understand the effects of interactions themselves as their de- sign target when exploring a design space”(Lim et al, 2007)

What designers explore with the idea of interaction gestalt is the space of emerging shapes of interactions; it is not about how interfaces look like or what features need to be implemented.(Lim et al, 2007)

It is important to understand that the attributes are not supposed to be used individually. As the original meaning of gestalt tells us, the sum is different from the whole. (Lim et al, 2007)

The Gestalt laws of perceptual organisation

Gestalt has produced laws of perception that describe the way people perceive patterns in visual images. Specifically, they propose means of organising visual elements effectively into structures (Chang et al, 2002). These laws provided designers with a powerful, comprehensible and scientific method to group design elements efficiently referring to the layout as a whole rather than its individual parts. Thus, the principles of Gestalt were adopted mostly in visual communication but also in other disciplines like architecture, human-computer interaction and linguistics (Graham, 2008). Furthermore, gestalt laws support designers in creating patterns to display data which facilitate user’s perception (Ware, 2003). Apart from supporting design decisions in terms of the aesthetics, gestalt principles also augment users’ cognitive processes (Fraher, 2010).

There were originally 114 laws introduced but most of them were very similar and overlapping. As a result only a few of them were practically meaningful and gained ground in the practical field. Most of these laws are described below (Carlson and Heth, 2010; Chang et al, 2002; Graham, 2008; Soegaard, 2010; Ware, 2003; Moore and Fitz, 1993; Wertheimer, 1938b):

- **The law of Prägnanz** is also mentioned as the law of good figure or simplicity because it is based on the idea that people tend to experience reality in a regular, simple and symmetric way. This occurs because people form only one interpretation per time by observing either the objects or their background and vice versa. Prägnanz is a fundamental notion in Gestalt theory on which all others laws are based.
- **The law of figure-ground** refers to the identification and the discrimination of objects with particular shape and location, namely ‘figures’, from their background. It is a fundamental law in gestalt theory which can be also referred with the term ‘contrast’. If it is applied efficiently, it makes elements like images or texts visible and easily understood by providing the proper visual feedback to the user. The issue of figure-ground is of great concern to the interface designers community and various guidelines have been established to help designers achieve it (World Wide Web Consortium, 2008). Graham (2008):
The law of proximity proposes that elements which are close to each other are perceived from the viewer as forming a group or at least as related, whereas elements within a big distance are understood as unrelated. Graham (2008):

In interactive media design, the closer items are spatially or temporally located near each other, the more likely they are to be considered part of an organized and unified group.

The law of closure states that people have the tendency to understand forms despite featured gaps or uncompleted lines. Especially when the form is familiar, people can identify it even if more gaps are added.

The law of similarity proposes that people have the tendency to group similar elements together. Similarity can be achieved in terms of shape, size, colour, proximity or direction. When applied to interface design, the law of similarity guides the user’s attention and helps elements categorisation.

The law of continuation is based on the fact that the human eye tends to follow a line or an array of elements, even if it is disrupted from other elements. If there are various possible routes, the eye tends to follow the simplest, smoothest and most predictable one.

The law of focal point. The focal point of any experience is the point, which attracts most attention and where emphasis is mostly given. The viewer is attracted from the focal point and follows the messages that start from it. Any designer puts effort to attract their users or viewers to the intended focal point. User tests with the eye tracking technology can prove if the focal point accomplishes its goal.

The law of balance and symmetry promote the notion that symmetrical and visually balanced objects are more easily perceived as a whole. Symmetry and balanced can be achieved in terms of shape, size and position. Unconnected but symmetrical elements can also form a group. However, unsymmetrical or unbalanced objects might be perceived as incomplete. A non-symmetrical and unbalanced design does not draw users attention and cannot guide them to a focal point.

Criticism & Limitations

- The gestaltists believed that perception came from innate processes, while the current view is that it is a combination of learnt and innate processes.
- People can view same figure in different ways (Quinlan & Wilton 1998).
- The gestalt view is descriptive rather than explanatory. No attempt is made to explain the processes within in the brain, and explanations tend to be post-hoc, so it is difficult to make predictions.
- It uses vague language. For example, it’s not clear what a good or simple shape is (Bruce, Georgeson and Green 2003).
- The approach is a bottom-up approach, while there is much evidence that top-down processes, based on prior knowledge, affects perception. This is discussed later.

7. Perceptions of groupings

(NOTE: Figure numbering is not continued from the start of the page and re-starts as ‘Figure 1.’ Will update this to the correct numbering as needed, once other diagrams make it into the chapter.)

The perception of grouping, also known as “perceptual segregation”, concerns the human ability to identify parts of visual information that belong together and thus form distinct objects. In addition to the Gestalt principles of proximity and similarity, object features and concepts of uniform connectedness, the configural superiority effect and figure-ground segregation influence how people perceive grouping. These are discussed later in the section.

The perception of visual information, closeness and similarity in the display will be taken to represent relatedness and, therefore, will facilitate perceptual grouping.

Features of objects also influence perceptual grouping. See Figure 1, below. The dotted line indicates how perceptual grouping should occur.

So what happens when these principles are in conflict?

Initially, humans will group objects by proximity and similarity (with a roughly equal proportion favouring each). However, if objects are already grouped by proximity, additional processes occur. If there are differences between the items within a cluster (e.g. shape, colour), then there is a within-cluster mismatch. The same concept can be applied between clusters: a between-cluster mismatch occurs when there are differences between groups, as well as within each one (See Figure 2).

In Situation A, there are no conflicts, so the principle of proximity will dominate.
In Situation B, the principles of proximity and similarity will prevail.
In Situation C, observers will choose either between the principle of proximity or the principle of similarity.
In Situation D, observers often ignore proximity and base their perceptual grouping on colour (above shape and texture).

Uniform connectedness refers to the tendency to group connected regions that share the same properties, e.g. colour, brightness, texture, and is often considered as a Gestalt principle in its own right. It has been found to dominate the principle of similarity (for all conditions) and proximity (when more than two objects are presented). This supports the suggestion that it occurs earlier in the perceptive process, although to what extent requires further investigation (Eysenck and Keane, 2010).
Most research in this area has been based on artificial figures, and further investigation is required to identify the extent to which these concepts can be applied to real-world situations.

The configural superiority effect describes how perception of grouping is affected more by organisation than complexity. In other words, it is easier to comprehend complex items that are organised clearly, than simple items that are organised badly. For designers, this emphasises the need to consider visual grouping and perception when communicating complex data (Pomerantz, 1981 in Eysenck and Keane, 2010).

Figure-ground segregation is “the perceptual organisation of the visual field into a figure (object of central interest) and a ground (less important background)” (Eysenck and Keane, 2010). Observers pay more attention to the main figure, thus perceiving it more clearly (facilitating improved recall) and less to the ground.

See the famous faces-goblet illusion for example, Figure 3. When the goblet is the object of interest - the figure - it appears in front of a white background. However, when the faces are the figure, they appear on black background (ground).

Speed of perceptual grouping is also influenced by:

- **Past experience**
  - familiarity with items will increase the speed with which they are processed

- **Domain knowledge**
  - an implicit understanding of the items may result in perceptual groupings that draw on aspects not explicitly shown in the visual information

**Application to design**

Applying these concepts to the design of visual displays will help facilitate observers’ understanding and comprehension, resulting in a more efficient, effective and pleasurable interfaces. Applying the concepts of perceptual grouping to the design of visual displays first and foremost requires knowledge of the users and their goals.

1. Identify the target user population and their tasks
2. Determine the key information required to complete tasks
3. Identify useful “groupings” to aid comprehension
4. Use the principles above to facilitate perceptual segregation, e.g. where possible:
   a. Group by proximity and/or similarity
   b. Connected related regions by making their features the same
   c. Avoid mismatches
   d. Draw attention to task-relevant information (figure) and eliminate irrelevant information (ground)

**8. Psychological pop-out**

Psychological pop out (also referred as visual pop outs) are characterised as visual objects that are the most salient in a display thereby grabbing visual attention.

Research in this area refer to the process of identifying pop outs as pre-attentive (parallel) processing whereby the entire visual field is scanned for basic features such as colour, contrast, line closure, line ends, contrast, tilt curvatures and size. (Treisman, 1986).

This process suggests that pop outs can be seen as a bottom up process (Woodfill & Zabih, 1990) where the pop out objects dominate attention which may then be processed cognitively depending on its feature relevance by the viewer.

Research gauging the processing of pop outs relies on measuring the reaction times (ms) of participants in identifying target pop out features. From the findings, there is consensus for pop out objects having basic features which are unique from objects in the same visual field.

These are very simple examples of object features that are scanned for pop outs, examination of the research reveal that saliency has to be rapidly distinguishable in the pre-attentive process otherwise serial processing occurs whereby each object in the visual field is scanned one-by-one, a slower process. (Treisman, 1986). Hence, while a red circle in an array of green circles will pop out, a red circle in a field of red squares and green circles will not as the features of colour and shape of the unique object are shared by other objects. (Krummenacher et al 2009).

Familiarity of objects assist in identifying pop outs more rapidly but only when familiar objects serve as distractors rather than target pop outs. (Wang, Cavanagh & Green, 1994). In their experiment an inverted N was more rapidly detected when looking for it in an array of normal N’s suggesting that familiar objects are grouped together in the visual process causing the odd object to stand out.

Luck and Hillyard (1994) found that if users expected target information in a pop out, this would further influence pop out detection.

**9. Icon structure and search**
10. Attention

With the vast amount of information in the world, perceived via all of our senses, it is important that we have a mechanism that prevents us from suffering from information overload. In other words, it is important to have a mechanism enabling us to optimise our limited informational processing resources towards stimuli that are relevant to our survival and goals. Attention is such a mechanism, which facilitates what we perceive and how we can act upon what we perceive. More specifically, attention enables us to reduce processing of irrelevant stimuli; enhance relevant stimuli; bind incoming information signals into coherent representations of the world; and recognise stimuli (Evans et al., 2011).

One of the earlier famous definitions of attention was given by the pioneering psychologist William James in 1890:

“Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalisation, concentration of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others.” (James, 1890, pp. 403-403)

This definition highlights the selective nature of attention. Selective attention may for instance occur when you search the item listing of an online shopping website for a specific household item. During tasks such as this, some pieces of information are registered, whereas other information is ignored.

What is excluded in James’ definition of attention, however, is the notion of divided attention. An example of divided attention is when you are trying to type up a coursework assignment while simultaneously chatting with a friend on Facebook. The limitations of the amount of incoming information you can process, i.e., your attentional processing capacity, is relevant to divided attention. See Multitasking.

James’ definition also focuses on attention as a conscious process; however, this notion has been qualified in later models. One such model is Schneider and Shiffrin’s automaticity model (Schneider & Shiffrin, 1977), which makes a distinction between controlled processing and automatic processing:

- Controlled processing is slow and conscious. Because such processing places substantial demands on an individual’s attentional resources, it is of limited capacity. An example of controlled processing is when you first learn to drive a car and need to focus all your attention on the skills and rules necessary to acquire to perform the task successfully.
- Automatic processing is fast and unconscious. Such processing does not make any demands on an individual’s attentional resources, and is therefore not constrained by capacity limitations. While controlled processing behaviour is easy to modify, automatic processing behaviour is not. An example of this is when you have become an experienced driver and are able to effortlessly retrieve all the skills and rules of driving from memory.

A common criticism of the automaticity model is the notion that automatic processing does not affect attentional resources. The Stroop effect is an infamous example of how attention directed towards a specific task can be detrimentally affected by automatically processed tasks: when given the task to name the colour a word has been written in, and the word spells out a conflicting colour (e.g., the word ‘blue’ written in the colour red), participants often struggle to correctly perform the task (Stroop, 1935).

11. Attention in different modalities

The various forms and processes of attention are underpinned by the sensory modalities. The modalities of vision, hearing and touch all contribute to our sense of attention individually and collectively. The exact way these modalities interact and inform each other is not completely understood (Eysenck & Keane, 2010). Below we will give an overview of the major modalities and how they operate, and we will look briefly at cross-modal attention.

Vision

Perhaps one of the most studied modalities for attention is that of vision. For focused visual attention there are two systems, one stimulus-driven and the other goal-directed. Although they utilize different parts of the brain, they interact and can influence each other. Thus, when engaged in a visual task, your visual system can become less responsive to visual distractions that are not task-related (Eysenck & Keane, 2010). Furthermore, the focus of attention can be location-based and object-based. Studies using fMRI have found brain activation in areas associated with spatial and object-based processing when presented with either spatial or object-based stimuli (O’Craven, Downing & Kanwisher, 1999). Finally, susceptibility to distraction is influenced by the perceptual load and executive cognitive control needed for the task. Perceptual load refers to the amount of stimuli and processing power needed for each stimulus. Executive cognitive control is concerned with higher functions such as memory. The more perceptual load needed for a task the less impact there is from distracting stimuli, however higher levels of executive cognitive control result in increasing impact of distractions (Lavie, 2005).

Audition

A previous editor has left these references here. This section should be expanded and the references should be moved to the references section at the bottom of the page - Sandy

Dichotic listening tests have been used to study attention in the auditory mode. Participants are simultaneously presented with two different auditory stimuli and then asked to distinguish between the two and recall the content of either message. Cherry (1953) used the selective attention experiment where the participant repeats aloud the content of one of the messages - known as shadowing. Cherry found that people recall the shadowed message poorly, suggesting that most of the processing necessary to shadow occurs in working memory and is not stored in the long-term memory.

There are 3 main theories explaining dichotic auditory processing (see figure 1):

Broadbent (1958)

Two simultaneous messages gain access at the same time to a sensory buffer. One of the inputs is then allowed through a filter on the basis of its physical characteristics, with the other input remaining in the buffer for later processing. This filter prevents overloading of the limited capacity mechanism beyond the filter, which processes the input thoroughly (e.g., in terms of its meaning). Criticism of this theory suggests it cannot account for variability in the amount of analysis of the non-shadowed message.

Treisman (1964)

Treisman’s attention theory shows the filter reduces or attenuates the analysis of unattended information. Treisman claimed that the location of the bottleneck was more flexible than Broadbent had suggested.

Deutsch and Deutsch (1963)

There is complete perceptual analysis of all stimuli so there should be no difference in detection rates between the two messages. Studies have since provided evidence against Deutsch and Deutsch’s theory.

More recent studies have supported both the Treisman and Broadbent approach although there are limitations of research in this area:

- it is very hard to control the onset and offset of auditory stimuli
- all three theories are vague making difficult to provide definitive tests of them
- finding out where selection takes places may not help to understand why or how this happens.

Somatosensory

The body’s somatosensory system processes several modalities which include:

- touch
- temperature
- proprioception (body position)
- nociception (pain).

As with audition and vision these senses are subject to “filtering” which can enhance sensitivity in areas of attention and filter out unwanted stimulus (Sambo and Forster).

Cross modal attention

Our sensory modalities do not operate independently of one another, in actual fact we use multiple senses at the same time. Cross-modal attention is the coordination of input from varying modalities (Driver & Spence). Lip reading is a good example of cross modal attention, where input from 2 modalities (in this case audition and vision) are co-ordinated. Cross-modal attention can occur when visual attention in a given location attracts auditory and/or tactile (touch-based) attention to the same location. Alternatively, directing auditory/tactile attention to a given location can attract visual attention to the same place (Eysenck & Keane).

When users expect a target in one modality to appear in a particular location, their judgements improve at that location - not only for the expected modality but also for other modalities (Driver & Spence). Interesting illusions can results when sensory modalities combine, as witnessed with the McGurk Effect. The McGurk effect is an illusion that combines auditory and visual perception. What you see influences what you hear. Syllables that sound similar or require similar movement in pronunciation can result in misinterpretation of the syllable being pronounced – for example the syllable ba/ga.

12. The influence of task/goals on visual search: the role of top-down processes

As opposed to interpretation emerging from observation, or bottom-up processing, top-down processing suggests that interpretation is influenced largely by knowledge, expectations, tasks and goals. This means that the context of the task and the goals of the individuals will have a strong influence on the processing of information and the cognitive approach to search.

According to Corbetta & Shulman (2002), top-down processing is the “flow of information from the ‘higher’ to ‘lower’ centres, conveying knowledge derived from previous experience rather than sensory stimulation”. They identified that, on a neurological level, different parts of the brain are utilised in activating top-down selection to that of the bottom-up system.

The influence of goals on attention has been demonstrated using eye tracking systems. One such study was carried out by Yarbus (1967); participants were initially asked to look freely at a painting, and eye saccades were mapped (fig. x1). Fig. x3 on the other hand, shows the focus of attention when participants were asked to give the ages of the people in the painting. Here the fixations are clearly on the faces of the people, and less on the environment in general.
13. Top-down and bottom-up control of attention

Attention is a selection mechanism in human mind which depends on the interaction between exogenous and endogenous factors (Posner, 1980).

**Definition of Bottom-up and Top-down**

There are two types of attention models. One is the bottom-up processing which makes the attention be driven by the salient visual properties of the objects. It is thought to operate on raw sensory input and does not depend on observer’s knowledge of the stimulus (Connor, Egeth & Yantis, 2004; Wolfe, Butcher, Lee & Hyle, 2003). For example, the flower in the picture is the most salient item because of its unique shape. However, if there is a task which required to find out other normal geometric shapes, e.g. the white square. The attention will be influenced.

The other one is top-down processing applying longer-term cognition and depending on observer’s knowledge (Connor, Egeth & Yantis, 2004; Wolfe, Butcher, Lee & Hyle, 2003). Human assigns the priority of attention to objects based on the task, behavioural goals or prior knowledge (Hornof and Halverson, 2003). As a result, the task strategy and internal goals determine the selection process of which object is relevant (Yantis, 1998). For example, there is a figure (Wolfe, Butcher, Lee & Hyle, 2003) where the attention will be driven towards the target’s property. Simultaneously, the most salient item will still draw some attention as well. The former is the top-down and the latter is the bottom-up.

**How to reduce the inference from Bottom-up**

Nevertheless, if some changes were made to figure 1.1. E.g. change the color of one of the white square to red. Then the inference from the bottom-up will be reduced when top-down is needed. Because there is a stronger difference between the target and distractor after that, it will contribute to the direct attention to the target and more efficient feature search (Wolfe, Butcher, Lee & Hyle, 2003). Besides the way above, there is another approach benefiting the reduction of bottom-up salience. It is to increase the diversity of the distractors. Actually, it also helps to reduce the saliency the most salient item.

**The comparison between Bottom-up and Top-down:**

Apparently, the top-down process is an intentional, deliberate, voluntary, effortful mental process, and it has a sustained time course (Yantis, 1998). On the other hand, the bottom-up process is an autonomous process, and it has a rapid time course (Posner et al., 1980). However, these two processes are not separated when human allocates the attention on objects. They are influencing and interacting with each other (Grossberg, Mingolla and Ross, 1994; Muller, Humphreys and Donnelly, 1994).

For instance, the famous researches of abrupt onsets showed that abrupt onsets draw attention even when there is no informative cues (Yantis and Jonides, 1984) and when the subjects were encouraged to ignore the cues (Jonides, 1981). At the same time, other researches also showed that although the task goals determined whether to ignore a singleton or not, other task-irrelevant dimensions automatically draw attention at the same time (Pashler, 1988).

We can take Fig. 1 as an example. Undoubtedly, most of people will pay attention to the full-blown flower in the middle at first sight when they see this picture, rather than notice the shape and quantity of leaves in the corner or the color of the metal baluster under the flower. Why the focus of attention of people like this? One reason is that the flower in the picture is “Salient”. Its color is orange, which high contracts the green background and makes itself standing out from the whole picture. In other word, this bright color is physical input and gives human visual stimulus and the bottom-up (stimulus-driven) control of attention leads people to look at this flower. In some situation, the flashy Ads can catch attention of people is on the same theory. Another possibility is that the flower is the biggest object in the centre of the picture, which will let people consider this flower as the target object of the photographer. So that viewers will think the full-blown flower as the main character and other things are the background of the picture. Under this cognition, people will pay more attention on main character rather that background. This is what we said Top-down (concept-driven) control of attention. In Figure 1, both bottom-up and top-down controls of attention focus on the same object, the flower, so we will look at it without conflict.

14. Application to systems

15. References


Gregory, R.L 1997 “Knowledge in Perception and Illusion”


Hassenzahl, M 2010, “Experience Design: Technology for All the Right Reasons”, Synthesis Lectures on Human-Centered Informatics,


16. Exam questions

Previous exam questions

1. Present an account of human attention, in terms of top-down and bottom-up processing, and of how multiple streams of information are processed during completion of dual tasks. Discuss the implications for the laws governing use of mobile phones whilst driving (2006).

2. It has been proposed that users plan their actions before they perform them. An alternative view is that action is situated, and that users behave by reacting to events as they happen. Discuss both perspectives, and consider what the implications of each are for the design of interactive systems (2007).

3. Why is it sometimes advantageous to present information to users using multiple modalities, and what are the difficulties of doing so? Include implications for the performance of both single and dual tasks in your answer (2007).

4. Present an account of human attention, in terms of top-down and bottom-up processing, and of how multiple streams of information are processed. Discuss the implications for the design of interactive systems, illustrating your discussion with examples (2008).

5. Discuss the strengths and limitations of various forms of representation (e.g, graphical and textual, declarative and procedural) for presenting different types of information to the user. Imagine you are designing an interactive ‘walk up and use’ information system for tourists to be located at strategic points around a city; sketch a few sample screens for such a system and present a design rationale for your main design decisions, focusing on how information is searched and presented (2008).

6. Describe how both visual layout and a user’s goals can influence visual search behaviour when searching for a link on a busy webpage. In your answer, explain how this knowledge can be used when designing such interfaces (2011).

Other exam questions that we’ve thought of...