Chapter 7

Learning to See

How are the edges, colours and distances interpreted as visual objects? The way the challenge is met is demonstrated by those ambiguous figures that turn up in so many works on visual perception. Consider this one, devised by Edgar Rubin in 1915. Is it the outline of a vase, or is it two faces? It can be seen as either - but only with much practice as both at once. Below is the even more venerable Necker cube. Probably you will see it first as a cube viewed from above. Look at it long enough and it will suddenly turn into a cube viewed from underneath. If you haven't met the Necker cube before it may take a while before your perception of it suddenly flips, but with practice you can switch back and forth between the two perspectives fairly easily.

This is curious. The retinal stimulation and the information transmitted to lateral geniculate and on to visual cortex remain unchanged and yet the conscious experience alters. We see a cube from below or a cube from above, we see two faces or a vase. Clearly the conscious experience is not simply a reflection of the retinal input. It's an interpretation of it, translating it into known objects. Somewhere between the registering of lines and edges and the conscious perception decisions have been made about just what the lines and edges mean - decisions that are usually quite unconscious.

There's also a third way of seeing the Necker cube, and there's significance in the fact that it's generally ignored. It could reasonably be interpreted simply as a two-dimensional pattern, not particularly meaningful. To someone brought up in wholly natural surroundings and ignorant of cubes it would be just that. And to a man who had lost his sight in infancy but had it partially restored by a corneal graft at the age of fifty, it was just a jumble of lines. But for those accustomed to a largely right-angled world, and schooled in geometry and its diagrams, making a three-dimensional interpretation is automatic.

The influence of experience on perception was extensively explored by Jer-ome Bruner and his colleagues. Using a tachistoscope - a machine which projects an image for a very brief and precisely measured period of time - they established that familiar stimulus patterns are 'seen' more quickly than less familiar ones, and less familiar ones more quickly than wholly unlikely ones. In one experiment, for instance, Bruner and Leo Postman showed brief exposures of nonsense words, and found that those which were constructed fairly like English words, in that the letters followed each other in combinations commonly found in English, were identified at significantly shorter exposures than those in which the sequence of letters was not characteristic of English. Expectations, they found, could also affect perception. When subjects were asked to identify words and were told what grammatical class of word to anticipate, they recognised those which conformed to the prediction at shorter exposures than those which didn't.

An experiment with playing cards underlined the influence of both past experience and current context. The trick was that while some of the playing cards flashed on the tachistoscope screen were conventionally coloured hearts and spades others were unorthodox - black hearts and red spades. Each card was shown first for ten milliseconds, and the exposure time then gradually increased until the subject could describe the card correctly. The average exposure required to identify the normal playing cards was 28 milliseconds. The average exposure needed before an unconventional one was identified was four times as long - 114 milliseconds. Moreover in about 10% of cases the incongruously coloured cards remained unrecognised even at the longest exposure used, a full second.

Clearly the familiar pattern was being converted into a conscious visual experience much more readily than the incongruous one. But a breakdown of the results showed that the expectations put in place by the first identification also played a part. When the first card presented to a subject was an unorthodox one the average exposure needed for a correct perception was 360 milliseconds. When the first encounter with an unconventional card followed one or more experiences of a standard card the average time required rose to 420 milliseconds. But after the subject had successfully identified one trick card the average exposure needed for the next one dropped to 230 milliseconds, and for a third or fourth it shrank still more substantially, to 84 milliseconds.

Particularly interesting is what the subjects reported before they got it right. The most frequent reaction was simply to see the unconventional card as a conventional one, some people being regularly influenced by colour, others by shape, while some were equally likely to make a judgement based on either. Another response was to describe the card as a conventional one, but with the colour distorted in some way - the descriptions included brown, purple, rust, olive, and black seen in red light. And some subjects, after reporting inaccurately but fairly confidently when the trick cards were seen at the briefest exposures, became totally confused as the exposure grew a little longer, and found it hard to make any sort of guess as to what they were seeing.

As a result of such experiments Bruner and Postman suggested a theory of expectancy, or, as Bruner would also call it, of perceptual readiness. Perception they said, involves an act of categorisation. And perceptual readiness *refers to the relative accessibility of categories to ... stimulus inputs. The more accessible a category, the less the stimulus input required for it to be sorted in terms of the category*. The neuropychologist Donald Hebb tackled the same idea using the word 'identity'. *The object that is perceived as having identity is capable of being associated readily with other objects or with some action, whereas the one that does not have identity is recalled with great difficulty or not at all, and is not recognised or named easily. Identity of course is a matter of degree and .... depends on a considerable degree of experience.*

For me both *category* and *identity* carry rather too much baggage from their everyday meanings. Things of very different appearance can share the same category, and identity is too often applied to the individual rather than the class. In company with many modern psychologists I prefer to follow Bartlett and Piaget and use the term *sensory schema*. A conscious experience, it seems, is generally created by matching current input to an established schema. If the sensory input adds up to a familiar stimulus pattern the matching is done immediately, and perception is effortless. If there's no appropriate schema a new one must be created, and achieving a clear perception takes longer.

An activity in which well-learned schemata play an obvious part is reading, and Bruner's concept of perceptual readiness has been nicely illustrated by studies of the way people read alphabetical languages. An experienced reader doesn't need to look at every letter, but skips along the line, fixating only on every fourth or fifth word. The longer, more unusual and most informative words are allotted to central vision, and the gaze slides over the little, thoroughly familiar ones - the conjunctions, articles and pronouns - identifying them without needing the sharpest focus. Recognising words involves paying particular attention to their initial letter, to their length, and to their overall shape. This can be deduced because when a mistake is made the incorrect word is usually one of roughly the right length and shape, and with the correct initial. The sensory input is matched to the wrong schema, but one that has similarities to the right one.

At the same time reading involves expectancies about meaning and grammar. A misread word is likely to be one that still makes sense in the context, and is the right part of speech. If, as the sentence continues, it turns out not to fit, in either respect, the reader is more likely to notice the mistake and go back to correct it. So reading is guided both by sensory schemata and by the expectations aroused by context. If the meaning or the grammar is very abstruse we have to read more slowly, and identify the individual words more carefully.

On the whole, though, reading can be a highly efficient process because the possible expectations are limited. There are only so many letters, the rules of the language only allow them to be combined in so many ways, the number of possible words, though very large, is not infinite, and we can expect sentences to say something we recognise as meaningful. Thus at every level there is a restricted number of schemata to be considered, along with a relatively modest number of possible hypotheses as to the interpretation of the data, so one doesn't need to register all the information available in order to construe the sentence. Thanks to these factors we can often make sense out of distorted or unconventional hand-writing. Though few of the individual letters are identifiable when considered separately, the assumption that they add up to known words, reasonable grammar, and some sort of meaning, helps in deciphering them. 'Reading' the natural visual world is a very similar task, but since the natural world is not constrained by so many predictable limitations it can sometimes be a demanding one.

A great deal of the learning that produces visual schemata is done in infancy and early childhood, but even in adulthood we sometimes encounter something novel - perhaps a complicated collection of edges and corners with a three-dimensional arrangement not readily revealed by binocular disparity or motion parallax. Or we come across a familiar object in misleading illumination. We stare at confusion, and then suddenly we realise what it is. Perhaps we run our fingers over the multiplicity of edges and check the ins and outs, or we learn what the machine is doing and how the dimly lit bits in the furthest corner must connect, or someone points out a distant shadow on the hills as the fold through which the road runs. And suddenly the shapes become clearer. It's almost as if one had been looking through a vaselined lens, and the lens has been wiped clean. The whole conscious experience is subtly modified as a decision is reached as to what one is seeing.

If the wrong schema is evoked at the beginning of the process, however, confusion results. Bruner and Mary Potter showed brief glimpses of photographs projected with varying degrees of lack of focus, and asked their subjects to identify the pictures. There were three levels of definition, lightly blurred, middling blurred and very blurred, and three exposure times, 122 seconds, 35 seconds and 13 seconds. The subjects were divided into nine groups, and each group first saw the pictures at one of the nine possible combinations of fuzziness and exposure time, and then progressed through other possible combinations.

The most striking result was that subjects who saw a picture first at a very blurred level of definition and a short exposure had great difficulty in working out what it represented when they saw it under more favourable conditions. The guesses made after a brief glimpse of a fuzzy and unidentifiable picture got in the way of accurate perceptions when they viewed it for longer, or with better focus. Subjects who viewed the pictures first under the more favourable conditions made far more successful identifications than those who saw it under these better conditions only after a first viewing had led them to make wrong guesses.

The essence of perception - or conscious sensory experience - was summed up by Hebb. "We do not really see a thing until we have recognised it."

What exactly is a visual schema?

I said in Chapter 2 that what we record is essentially a pattern of relationships, and gave as an example the visual concept, as I called it there, of a cow. A cow is a collection of components joined together in a consistent way, such that the spatial relationships between some features can change, within certain limits, and that between others cannot - unless perhaps it's dead or damaged. That definition would fit most vertebrates of course. What distinguishes the various species is essentially the shapes of the various components, and the degree to which the spatial relationships can vary. Colour and size are also important, expected to vary only within certain limits. Building a schema is well described in Eleanor Gibson's words - a process of *extracting the invariants in stimulus information*. It means learning which aspects of a stimulus pattern remain consistent in all its manifestations, and which differences serve to distinguish it from other, similar patterns. A preliminary guess at an identification can then be confirmed by directing attention appropriately.

A sensory schema must always, of course, be capable of modification in the light of new experience. Piaget's account of infant learning has proved to fit extremely well to what has been discovered about the learning involved in perception. To use his terms, new information must frequently be assimilated into an existing schema, and the schema must be adapted to accommodate it.

The essence of a schema is its functionality. We need to recognise visual objects in order to know what to expect of them, and what we might need or want to do about them. Just how much detail is necessary in order to make useful predictions about a visual stimulus pattern can vary considerably, depending both upon the nature of the pattern and on the needs and interests of the observer. Schemata can therefore differ considerably in their precision and the amount of detail recorded, and in the degree to which a general schema is divided into more closely determined subcategories. What is just another brown horse to many eyes can to one perceiver be quite clearly my horse Dobbin. A coin may be just a penny or an interesting piece of metallic art. The exactitude of the schema depends on the amount of attention paid to the exemplars.

A visual schema has to be very clear about the essentials of the class of percepts it represents. But at the same time it must usually be fairly rough-hewn, for in most cases it has to embrace quite a range of patterns of retinal stimulation. Dobbin is recognisable in many different poses, and from many different perspectives. Even something as simple as a geometric figure, a square or a triangle, can be classified regardless of its size, and is still seen as a square or a triangle when viewed obliquely. Schemata which cover whole classes of stimulus pattern must be flexible enough to be fitted to a pattern of visual stimulation which qualifies as a novel example of the familiar class, or to a familiar object in an unfamiliar orientation, or to one which is transformed in some considerable but essentially unimportant way. It's not surprising that while we are pretty good at recognising visual objects we are rather bad, for the most part, at dredging up accurate representations of them out of memory, so that even the most intimately known faces and scenes often cannot be recalled in full detail.

Distinguishing between faces is a particularly tricky business, since it involves making fine distinctions between many basically similar stimulus-paterns, and also discounting the effect of constantly changing expressions, and in the longer term the effects of passing years. This makes individual faces especially hard to remember exactly, so photographs of our friends are highly valued, and the portraitist who can capture what seems to many to be the essence of a face is the greatest success. The artist most relevant to this theme, however, may be the caricaturist, who aims not so much at portraying the actual face as at summarising the features which most distinguish it from other faces.

If the conscious visual experience is created with reference to stored knowledge, or schemata, the reverse must also be true. The visual schemata are forged through conscious experience. It's the process of turning the sensory evidence into a meaningful perception that creates the memory trace which is a schema.

An experiment by Deborah Chambers and Daniel Reisberg helped to confirm this, with ambiguous drawings like this one. Some people identified this as a rabbit, some as a duck, but few recognised both possibilities without prompting. The subjects were asked to memorise the figure, without their attention being drawn to its ambiguity. The drawing was then removed and the subject was shown another ambiguous figure, and this time it was pointed out that it could be seen in two ways. Then the subject was asked to think about the first drawing and say whether that too had a second interpretation. None of the subjects could re-interpret their remembered image, even after further bouts of coaching and some hint-dropping. On the other hand they could draw the figure from memory, and having drawn it could recognise its ambiguity. This supports the conclusion indicated by earlier evidence - what's stored is an interpretation of sensory input rather than a record of the input itself. It can be deduced that the work which produces perception is also the work which creates a memory trace. The conscious experience which is our sensory model of the world-out-there and the schemata which we derive from it are inextricably interdependent.

Infant learning

If experience plays such a big part in seeing, it follows that infants must have a great deal of learning to do before they can make sense of the world. They need to build up a visual vocabulary, as it were. The nature of infant interactions with the visual world was explored by Tom Bower. He discovered that the young infant treats something that stands out from the background by virtue of distance or movement as a single object. Show her a matchbox resting on a book and she will reach for book-plus-matchbox as if it were all one thing. Only when books and matchboxes have become familiar as separate objects does this behaviour change. And when Bower contrived to show infants a visual object which moved about as one unit for a while, before breaking up into several pieces which moved off in different directions, they looked very surprised. Clearly we have to learn by experience what visual arrays represent several separate objects, and are likely to be capable of coming apart.

Bower also found that young infants can be insensitive to the appearance of visual objects, but very sensitive to the speed of visual movement. When a moving object disappears behind a screen infants as young as eight weeks anticipate its reappearance very efficiently, turning their eyes to meet the exit point just as, or just before, the object emerges. Bower arranged things so that one object went behind the screen and a completely different one emerged at the other side after the appropriate interval, and infants of up to about twenty weeks continued to track this as if nothing had happened, unaffected by major changes in the shape, size and colour. Only older infants showed surprise at the transformation. (Perhaps this indifference to varying appearance is not quite as surprising as it seems if one considers that the most compelling visual stimulus patterns, people, tend to turn up differently dressed, often in quite different colours, from day to day.)

However, when an object disappeared behind the screen and an identical object appeared on the other side, but much sooner than would be expected if the original had continued at the same speed, the infants' eyes moved rapidly to the new object, went back to the point of disappearance, then jumped back to the new object again. And younger infants seemed to find this experience disturbing, and refused to look at the oddly behaving apparition more than once or twice.

Bower repeated the experiment in a new form, using an arrangement of half-silvered mirrors and cunning lighting, so that one object seemed to be trans-formed into another in full view. Infants of up to about four months still reacted in the same way. As long as the object continued on the same path at a constant rate, they continued to track it unconcernedly. But if the rate of movement altered the infants got upset. It was as if these very young babies *define the identity of a moving object solely in terms of movement*.

The next step was to find out what happened when a moving object simply stopped, in full view of the infant. It turned out that the gaze continued to travel along the path the object had been following, and the now stationary object was ignored. This might merely indicate that very young infants don’t have very precise control over eye and head movements and cannot stop tracking exactly on cue. But frame-by-frame analysis of videotape records showed that there was in fact a pause when the object stopped, and the eye movement only continued after a few tenths of a second.

Perhaps, Bower thought, the young infant doesn’t identify the stationary object with the moving one. If so, the reverse should also be true, and a stationary object which starts to move won’t be identified as the same object. This idea was tested with the aid of a splendid toy train with flashing lights. It stood in one place for ten seconds, moved to another place, stopped there for ten seconds, then returned to its starting point, and this shuttling sequence was repeated several times. Infants of between twelve and sixteen weeks of age at first went on with their tracking movement after the train had stopped, but after a few trials they settled down to follow it from A to B and back, without any overshooting.

Is the infant really tracking a single object from one place to another, Bower wondered. Or has she learnt to expect a succession of separate events - a stationary object at A, which disappears and is replaced by a movement from A to B at a certain rate, then a stationary object at B, succeeded by a movement from B to A? If the second account were the true one, then the infant would be taken by surprise if the object moved off on a new track and stopped in a new location. And so it turned out. When, after trundling from A to B and back again several times, the train went off in a new direction and stopped in full view but at a new spot, the infant would ignore it, and look firmly towards the accustomed stopping location, often looking surprised and puzzled when it did not see the train there. This seems to tie in with an observation of Piaget's. He noted how his infant son's gaze followed his mother as she left the room, and then turned back to where she had been standing beside his cradle, as if he expected to see her still there.

Bower's experiments showed how little the newborn infant knows about visual objects. They suggest that hardwired elements of the visual system, such as those V1 cells which are connected up so as to be sensitive to different directions and speeds of movement, must shape the infant's earliest responses. The first expectations would seem to grow out of inductive rule-making, or conditioned learning - what has happened several times can be expected to continue happening - and are only modified as visual experience increases and more complex rules about the likely behaviour of perceived objects are acquired. No wonder the very young regard the world with such a solemn, studious gaze.

More recently electroencephalography (which records the electrical potentials on the brain's surface with external electrodes) has shown that the infant capacity for registering visual motion does indeed mature before the capacity for registering visual form - or at least most sorts of visual form. Furthermore, a study of three individuals who had been endowed with sight after being blind from birth or infancy revealed that to begin with they too could only distinguish object from background when the one moved in relation to the other. It took them between ten and eighteen months to learn to use other cues.

For infants, who can concentrate all their efforts on investigating the world around them, the learning is easier. By six months they can cope quite efficiently with objects that change their trajectory while obscured from view. Kochukhova and Gredebäck confirmed that babies of this age initially anticipate that when an object travelling in a straight line disappears behind a screen it will continue on the same course to emerge on the other side. But after two or three experiences of an object which reappears instead at a right angle to its original path they learn to look for it there, showing no signs of distress at the failure to conform with their original expectation.

In addition to developing schemata for visual objects, there are other things to learn. To begin with, we must learn how to look. The fovea, that central area of the retina which is most densely packed with receptors and capable of fine resolution, is very small, covering only about two degrees of arc. In order to study any but the smallest of objects in any detail, therefore, the gaze must move over it, focussing now on one part, now on another. The movements are known as saccades, and in adults take up about 10% of the viewing time, occurring at the rate of two to four per second and often coinciding conveniently with blinks. The fixations are usually concentrated on denser clusters of lines or edges, or conspicuous angles, areas likely to carry the most information.

The adult style of scanning takes a long time to develop, however. Newborn infants hardly scan at all; they tend to focus on just one small point, with minor excursions around it, as if just that one bit of the visual scene was enough to occupy their attention span. By two months the saccades grow rather longer. But youngsters of three or four years old still spend a lot of time focussing on the centre of a complex scene, rather than tracking round it. Over the next two or three years the habit of concentrating fixations around the outline of figures becomes established, but there are still a great many small saccades around points of detail, and the features which receive such attention are often not those which an adult would consider most informative. Finally, with increasing experience, fewer saccades are needed, fixations become briefer, and there are longer traverses between fixations.

If infants concentrate their attention on one small cluster of features at a time it may well be because some part of the neuronal machinery has to mature before information gathered from a wider field can be meaningfully assembled. But the evidence also indicates that we must work out the most profitable ways of scanning a visual object, so as to derive the maximum of information with the minimum of effort.

In adulthood each individual seems to develop their own particular scanpath for a scene, and may repeat it several times. There will be irregular excursions outside this well-used pathway, but a fairly consistent re-tracing of one pattern of saccades and fixations takes up a significant proportion of viewing time. Moreover if the same picture is shown again, a little while later, the same pattern reappears, accounting for perhaps half the viewing time. The fixations are no doubt concentrated largely on those features of the object which are essential for making an identification, or on distinguishing the current example of the class. But it seems likely that a schema is somewhat influenced by the shape of the first scanpath. If an element of pure chance helps to determine how the attention is first drawn around an object, that could be why the memories of different individuals often emphasise different aspects of the same stimulus pattern.

As well as discovering the best technique for studying an object we also have to learn some general rules about interpreting the patterns of retinal stimulation. We must learn to sort out shadows, and to appreciate their significance. We find out how visual textures relate to tactile textures, and the significance of a gleam of pure white light reflected from an otherwise coloured surface. We learn about transparency, and about the way things seen through water are slightly displaced. We work out how coloured glass affects the colour of things seen through it. We discover the properties of mirrors, and the distorting effect caused by curved mirrors, or by transparent glass of an inconsistent thickness. All this acquired knowledge underlies our adult perception of the world.

However, there do seem to be one or two important things that don't have to be learnt. One is the effect of distance on the retinal size of an object. Bower established that, trained to respond differently to a large square and a small one, an infant doesn’t confuse a large square at a distance with a small one close up, even though the retinal image is the same in both cases. Equally, the infant can tell the difference between a square viewed on the slant and a trapezium viewed straight on. The inbuilt distance-measuring mechanisms serve to produce some basic assumptions about visual objects, it seems, just as the movement-detecting mechanisms do, and these assumptions point the infant in the right direction.

Occasionally it happens that someone who has been blind from birth has an operation which renders their eyes functional. They are then in very much the same situation as the young infant in trying to make sense of a visual world - but with the disadvantage that their expectations are likely to have been shaped by what they have learnt to achieve with their other senses, and they are not tackling the job with the infant's unprejudiced sense of adventure. Reports on just how much difficulty they encounter vary somewhat.

According to J.Z. Young, such a patient, opening his eyes upon the world for the first time *gets little or no enjoyment; indeed, he finds the experience painful. He reports only a spinning mass of lights and colours. He proves to be quite unable to pick out objects by sight, to recognise what they are, or to name them.* It takes a long time to learn to recognise visual shapes, and at the beginning it requires conscious concentration. *Of course, if I look carefully I see that there are three sharp turns at the edge of one patch of light and four on the other*, said a newly sighted man when confronted with a triangle and a square.

However, shapes for which there are tactile schemata may be recognised more readily. A subject studied by Richard Gregory and Jean Wallace immediately identified upper case letters, which he had learnt to read by touch, but took many months to learn lower case letters. He could tell the time by a clock too, since he had been accustomed to using a glassless watch. He could use his sight to make his way about without bumping into things, and he could judge the distance of objects he was already acquainted with by touch, but was bad at distance judgements otherwise. Many of the visual illusions experienced by the normally sighted (including the Necker cube) did not work for him, or operated much less strongly. Unfortunately there doesn’t seem to have been any follow-up to discover what improvement a few years of practice made.

A more recent case was a man born without lenses in his eyes, who got his first pair of glasses at the age of twenty nine. Their assistance still left him with a far from perfect retinal image, but over the first year he learnt to recognise simple, two-dimensional shapes, and in the second year began to cope with more complex objects. He reported that at first he saw a cow as separate patches of black and white, but perceived it as all one object once it began to move, and thereafter recognised it even when it was still.

Youngsters seem to do better, judging by five subjects, aged between eight and seventeen years, who were treated for dense congenital cataracts or corneal opacity. Though previously able only to distinguish between light and darkness, and in two cases to see a bright moving light, within 48 hours of their surgery they all performed pretty well on a visual match-to-sample task, using well defined objects constructed especially for the purpose. They were also successful at matching objects by touch. But at this point they were very bad at matching a visual shape to a felt one. Three of the subjects had improved significantly on this last task, however, when tested a few days later. In another study a girl of 44 months made rapid progress in integrating tactile and visual inputs.

Illusions created by learning

Creating schemata with which to interpret sensory input is a neat way of dealing with huge amounts of data, and doing it speedily. There's a record of an archetypal form, in which some features and the relationships between them are essential and others less so. If the sensory input fits the essential pattern it can be classified swiftly. Furthermore the classification can often be made successfully even when the data is ambiguous, or when the object is only partially visible. An example of the readiness with which we do it is provided by figures such as this one (devised by Gaetano Kanisza). We are so used to seeing objects partially concealed by other objects that the interpretation is automatic.

The readiness with which we evoke well-used schemata leads to all sorts of other illusions. Many are concerned with distance and size. Our inbuilt systems for judging distance - binocular parallax for nearer objects and motion parallax for further ones - are pretty effective. But as our knowledge of the world develops we learn additional ways of assessing distance. One of the characteristics that goes into the schema for a visual object is its likely size. Using this knowledge, and our appreciation of how one object can mask another, and of perspective, we have no difficulty in translating a two-dimensional drawing into a three-dimensional scene. A painting may even look more three-dimensional when viewed with only one eye, no doubt because the effect of learned cues is not contradicted by the absence of binocular disparity. For everyone accustomed to the use of perspective in the visual arts the habit of converting two-dimensional drawings into three-dimensional percepts leads to some well-known illusions.

Psychologists have had great fun devising three-dimensional arrangements which mislead the viewer - at least if only one eye is used - by not conforming to well-established expectations. One such trick array is shown here. The subjects, observing it through a carefully positioned hole, were asked to indicate the distance of the three surfaces by aligning a rod with each in turn. They generally interpreted the display as one would the flat reproduction. But in fact the smallest square was nearest, and properly speaking neither it nor the next nearest shape was a square at all, since corners had been cut out. The accessible schema easily triumphed over the less familiar.

Jolliest and most bizarre of the illusions of this type is the one created by Adelbert Ames, an artist turned psychologist. He designed and built a room which, when viewed monocularly from the appropriate spot, presented exactly the retinal image that would be produced by a normal, right-angled room. In fact - you will have guessed it - it was nothing of the sort. One end of the back wall was much further away than the other. But all the features which would normally supply perspective cues - the angles between walls, ceiling and floor, the shapes of windows and doors - were adjusted so as to create the same image as a conventional room.

Ames' rather startling discovery was that when someone looked into this distorted room and saw two people, one in each of the two further corners, and one therefore presenting a significantly smaller retinal image than the other, what was usually seen was a normal room containing two people of very different size. The assumption that the room was a conventional shape was the prevailing influence on the perceptual experience. We are conditioned to expect rooms to have right-angled corners, and human beings do vary in size quite considerably, so it's perhaps natural that the latter schema should be the more flexible one. The illusion doesn't work when the people in the room are known to the observer. Nor does it work on people whose perceptions haven't been conditioned by straight-lined, right-angled architecture.

How schemata are stored

The experiments outlined above mostly date back to the nineteen fifties and sixties. It seemed reasonable to guess that there might be areas in the brain where neurons could be regarded as functioning to record schemata. This would provide a way of explaining why damage in certain areas of the brain, mostly in the temporal cortex, can result in an inability to recognise a certain class of visual percept, while leaving the sufferer perfectly capable of dealing with other types. A stroke that affects one particular area, for instance, creates great difficulty with faces - the patient may be unable to recognise even their nearest and dearest, or their own face in the mirror. Damage to another area interferes with the ability to read, and a third area is important for recognising tools.

The anatomical evidence was consistent with the idea that the relevant area might hold the schemata for these particular types of visual percept, since after the retinal input has been filtered through the various stages of the occipital cortex the results are passed forward through the lower part of the temporal cortex. It wasn't until the eighties and nineties, however, that neurons towards the front of the temporal cortex of monkeys were identified with activity which suggested they might be the physical machinery embodying the hypothesised schemata.

The first discovery was of neurons that were activated specifically by monkey faces. Some fired for any conspecific face, others only for the face of a particular individual. It seemed a reasonable conclusion that the first sort corresponded to the generalised visual schema for a monkey face, the second to a more particularised schema. Some were activated by a full-face view, while others specialised in side-views; the degree of fussiness about viewpoint seemed to vary.

Then, in the early nineties, Keiji Tanaka found cells at the occipital end of inferior temporal cortex which were responsive to more modest assemblages of features. He and his colleagues carried out their explorations by presenting three-dimensional representations of plants and animals to his monkey subjects. When they found one that activated a neuron they removed different features from the image in turn, until they found out just what combination was sufficient to produce a strong firing rate.

Later work showed that the neat columnar organisation already seen in occipital cortex continues here. Within one column cells tend to respond to roughly similar feature arrangements. An adjacent column may deal in variations on a related pattern. No doubt the vast variety of potential inputs is dealt with here, as in primary visual cortex, by population coding - most inputs will excite more than one neuron, and it will be the precise combination of excitations that conveys the information to the next stage of the pathway.

Damage in this part of inferotemporal cortex causes difficulties with shapes generally, rather than with any particular type of percept. Receiving a major part of its input from V4, it might be described as assembling fragments of pattern into somewhat larger portions. Another contributor to this process is an area between V4 and inferotemporal, the lateral occipital cortex. Neurons here are sensitive to visual objects but indifferent to textures, don't care about the size of the object, and prefer fuzzy representations to clearly detailed ones. An fMRI study showed that when people studied novel shapes and learnt to identify them activity here declined if the shapes were easily distinguished from the background, but not if they were camouflaged by a similar background. All this suggests that lateral occipital is concerned with sorting out just what edges can be interpreted as adding up to a visual object, and what’s to be regarded as background - presumably with the help of distance and movement information. If the objects are already familiar, and easy to detect, it doesn't have to work very hard. The detail within the object, meanwhile, is left to later stages of the pathway to deal with.

At early stages of the pathway retinal position is still important, as might be expected since the stimulus patterns to which neurons are sensitive are generally only components of whole visual objects. Brincat and Connor carried out a study using a huge collection of subtly varied combinations of straight lines, curves and angles, and uncovered effects of relative position as well as shape.

As the visual pathway unfolds and the neurons focus increasingly on whole shapes and objects rather than lines and edges, so they have increasingly large receptive fields, with spreading dendrites that offer the potential for an ever larger number of synaptic inputs. And more and more of the receptive fields include the fovea.

From posterior inferotemporal cortex the results are relayed to the more anterior part, to be turned into information about whole objects, as in the face area. When subtle discriminations must be made among visual stimulus-patterns, with attention to detail, the neurons representing them don't fire more strongly as a result, but they become more finely tuned, and less likely to fire for stimulus patterns that only approximate the perfect one. Sigala and Logothetis created two series of drawings, one of schematic faces in which the positions of the features varied, the other of outline fish which differed in the shape of nose, fins or tail. Each series was composed of ten examples, divided into two categories, and while there were four variable features within each category only two of them were relevant to the task. Two monkeys were trained to pull one lever when they were presented with a member of one category, and a different lever when they saw an example of the other category, a discrimination which they learned to perform extremely well. Recordings were made from 150 neurons in the anterior inferotemporal cortex, of which 96 proved responsive to line drawings, including the test patterns, and to some natural images. Of these, 44 responded preferentially to a face which contained one particular version of a feature, or sometimes particular versions of two features. A very similar proportion of the 65 neurons that were also tested with the fish stimuli performed in the same way. The interesting discovery was that by the time the monkeys had become adept at their task (they achieved a 98% accuracy rate) the neurons which responded to the diagnostic features had become significantly more selective than those which responded to the irrelevant ones.

In a similar experiment Chris Baker and colleagues created stimulus patterns which were based on a vertical rectangle, or stem, with various different shapes attached at either end, to form something they called a baton. Monkeys were trained to distinguish eight of these batons altogether, again being rewarded if they pulled a righthand lever for some, a lefthand lever for the others. Then, for control purposes, novel versions of the same sort of stimulus were presented. Neurons which had developed a strong response to the learned stimuli no longer responded to other variants on the theme. All this suggests that inhibitory influences become more powerful as learning advances, preventing the cell from firing for mere approximations to a preferred stimulus pattern.

A rather different sort of experiment by Adam Messinger and colleagues confirmed that significance can be as important as appearance to inferotemporal neurons. Here the monkeys learnt, by trial and error, that four picture stimuli should be grouped into two pairs. First, having fixated on a small spot, they were shown a single picture. This was followed by a presentation of two other pictures from the set, and they were rewarded for looking at the one that was designated its pair. As their learning progressed the stimulus patterns that were paired came to elicit increasingly similar neuronal responses.

Presumably it's by similar means that we learn that Mother is always Mother, no matter how she's dressed, and Auntie is not Mother, even if she has borrowed one of Mother's outfits. What's learnt is where to pay attention to make a reliable categorisation, and this is what the schema-areas record.

The neuronal excitations in early stages of the visual pathway correlate to a fair degree with the pattern of retinal excitation. Those in the later stages tend to correlate with the conscious perception. Or, to be accurate, some ingenious experiments have shown that they correlate with what a monkey reports that it sees, and it seems reasonable to suppose that what it reports is a conscious perception. The experiments make use of ambiguous stimulus patterns that can be seen in more than one way, and the monkeys are first trained to make one response to one of the possible percepts, another to the alternative.

Leopold and Logothetis studied what happened when different images were presented to the two eyes. Humans in this situation see first one image and then the other, an unwilled alternation that usually takes place every few seconds. Monkeys seem to have the same way of dealing with the competing patterns of excitation. In inferotemporal cortex practically all the neurons the experimenters sampled changed their firing rate just before the monkey reported an altered perception. Earlier in the pathway the activity of about half the cells changed as the perceptions switched, and even in primary visual cortex about 18% modified their response.

Bradley, Chang and Andersen used a different illusion. It's possible to create on a computer screen a pattern of moving dots which humans see as a rotating transparent cylinder. It's still interpreted this way if about half the dots are travelling from left to right, and half in the opposite direction, but a cylinder seen as turning towards the right will suddenly convert into one that turns to the left. Monkeys were trained to signal by an eye movement which way a real transparent cylinder was revolving. Then they were put in front of the illusory cylinder on the computer screen while the scientists studied the activity in MT, the area in the middle of the temporal lobe that deals with visual motion. About half the neurons here varied their activity in tandem with the experience reported by the monkey. Cells that were normally activated by left-to-right movement would fire strongly when the monkey indicated it saw rightward movement, and much less strongly when the cylinder was seen rotating to the left - despite the fact that, as with the Necker cube, the retinal input was the same in both cases.

I hope it will be apparent by now that our conscious visual experience is a thoroughly creative interpretation of retinal input. It might be described as a guess as to what the retinal input is likely to represent. The guess is heavily influenced by acquired knowledge of the world, and sometimes it means that we see what we expect to see. It can also be influenced by the current context - Jerome Bruner showed that a line and an adjacent squiggle is interpreted as a B when it's surrounded by letters, as 13 when it's surrounded by numbers. And sometimes the conscious experience is coloured by current pre-occupations and desires, so that we think we see what we're looking for, and find we were mistaken only when we look more closely. As Richard Gregory put it: *normal everyday perceptions are not selections of reality but are rather imaginative reconstructions.*

In view of its creative nature it seems quite appropriate to refer to the conscious visual experience as 'the picture in the brain'. When first used this term was roundly criticised as implying the existence of a little homunculus sitting in the brain looking at the picture. I don't believe modern readers are likely to put this interpretation on the phrase. But if you do feel tempted that way, gentle reader, please don't imagine the homunculus sitting and looking. Conceive of him as painting the picture. He’s dashing about in manic haste, slapping on some paint here, filling in another bit of canvas there, rushing over to a distant corner to make a convincing interpretation of the signals he's received about that bit of the scene, and then haring back to the first area to update his work as fresh data come in. Then you turn your head, and he has to start all over again. (Terry Pratchett fans may like to think of him as performing a job rather like that of the elf inside the Discworld camera - but having to work considerably faster.)

The creativity that goes into the picture-in-the-brain is demonstrated by one particularly flamboyant bit of invention. The axons of the retinal ganglion cells all run to one spot on the noseward side, where they join together to form the optic nerve. Since the space is entirely occupied by axons there are no receptors here, and there is a small blind spot. This doesn't normally matter - the two blind spots cover non-equivalent bits of retinal field, and what is missed in one eye will be recorded in the other. But if you view the world with one eye closed there is no indication that something is missing - the brain simply fills in the gap to match the surround. You can prove this to yourself by making two conspicuous dots on a blank sheet of paper, about an inch apart. Hold the paper out at arm's length, close your left eye, and focus your right eye firmly on the lefthand dot. Then bring the paper slowly towards your nose without shifting your gaze. At a certain distance the second black spot will vanish, to be replaced by unsullied background.

The same filling-in process occurs when blind spots result from damage to the retina. This is unfortunate for people who develop glaucoma, a disease which destroys retinal cells, for they may not notice the gaps appearing in their vision until fairly significant bits of the visual scene begin to vanish, by which time the damage to the retina is well advanced. Most oddly, though, the creative cover-up can happen when the retina is functioning perfectly well and there is a genuine lack of visual input. If, in the far periphery of the retinal field, there is a texture that looks as if it should be a continuous pattern, but which actually has a small blank gap in it, the gap may quite rapidly vanish. That busy little homunculus with the paintbrush decides it's unlikely, and opts to override the data and cover it up.

The process of creating the picture might be described as a sort of proto-reasoning. It's done, most of the time, without conscious thought. The rules about interpreting the input are put to use, and the inputs are matched to the learnt schemata before the conscious experience even materialises. Even when a novel or imperfectly defined visual object is encountered the process of considering the evidence and clarifying the perception only rarely feels like something that should qualify as thought. In effect, the process must involve dealing with questions like: *is that a shadow or an edge? should greater weight be given to this feature or to that?* But the questions are not generally formulated in consciousness. Nevertheless, a juggling with learned concepts is involved in the creation of the conscious experience. The evolution of such sensory reasoning can be seen as an important step on the path towards abstract reasoning.

Recapitulation

Visual illusions show that our conscious visual perception is usually an interpretation of the retinal input, one that is greatly influenced by past experience.

Experience enables us to build up visual schemata, which are embodied in the connections of more forward areas of a pathway through the lower part of temporal cortex.

Such connections are readily activated, while visual inputs that don’t correspond to existing schemata take longer to be converted into clear perceptions, or are misinterpreted.

A schema can be defined as a record of an essential pattern of relationships, and of where attention should be directed.

The firing of neurons in the schema-areas is heavily influenced by the significance of a stimulus pattern. Details may influence the neuronal response only if they are important for distinguishing the pattern from others of a different import.

This is why we are good at identifying visual objects in varying manifestations, bad at recalling their exact appearance.

There’s a lot to learn before a really useful conscious visual experience can be produced. As well as building up a good collection of visual schemata, which includes knowing how to scan the most informative bits of the scene, there are all sorts of rules to master - about the effects of the direction from which light falls, for instance, and about matters such as transparency and reflection.

The neuronal activity involved in creating the conscious experience seems to be what creates the schema.

It has a lot in common with more abstract reasoning, and can be regarded as a forerunner of it.