VISUAL THINKING *for* DESIGN

Colin Ware

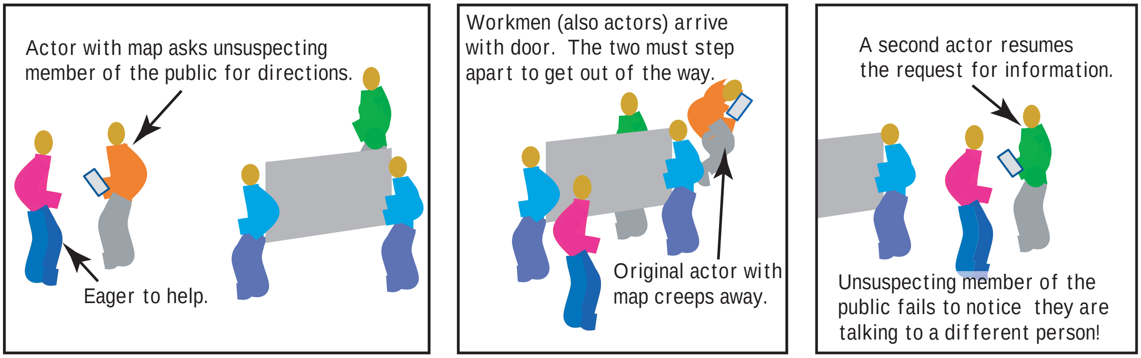
Chapter 1

Visual Queries

When we are awake, with our eyes open, we have the impression that we see the world vividly, completely, and in detail. But this impression is dead wrong. As scientists have devised increasingly elaborate tests to find out what is stored in the brain about the state of the visual world at any instant, the same answer has come back again and again—at any given instant, we apprehend only a tiny amount of the information in our surroundings, but it is usually just the right information to carry us through the task of the moment.

We cannot even remember new faces unless we are specifically paying attention. Consider the following remarkable "real world" experiment carried out by psychologists Daniel Simons and Daniel Levin.◆ A trained actor approached an unsuspecting member of the public, map in hand and in a crowded place with lots of pedestrian traffic, and began to ask for directions. Then, by means of a clever maneuver involving two workmen and a door, a second actor replaced the first in the middle of the conversation.

◆Daniel J. Simons and Daniel T. Levin. 1998. Failure to detect changes to people during a real world interaction. *Psychonomic Bulletin and Review*. 5: 644–669.



The second actor could have different clothing and different hair color, yet more than 50 percent of the time the unsuspecting participants failed to notice the substitution. Incredibly, people even failed to notice a change in gender! In some of the experiments, a male actor started the dialogue and a female actor was substituted under the cover of the two workmen with the door, but still most people failed to spot the switch.

What is going on here? On the one hand, we have a subjective impression of being aware of everything, on the other hand, it seems, we see very little. How can this extraordinary finding be reconciled with our vivid impression that we see the whole visual environment? The solution, as psychologist Kevin O'Regan◆ puts it, is that "The world is its own memory." We see very little at any given instant, but we can sample any part of our visual environment so rapidly with swift eye movement, that we think we have all of it at once in our consciousness experience. We get what we need, when we need it. The reason why the unwitting participants in Simons and Levin's experiment failed to notice the changeover was that they were doing their best to concentrate on the map, and although they had undoubtedly glanced at the face of the person holding it, that information was not critical and was not retained. We have very little attentional capacity, and information unrelated to our current task is quickly replaced with something we need right now.

There is a very general lesson here about seeing and cognition. The brain, like all biological systems, has become optimized over millennia of evolution. Brains have a very high level of energy consumption and must be kept as small as possible, or our heads would topple us over. Keeping a copy of the world in our brains would be a huge waste of cognitive resources and completely unnecessary. It is much more efficient to have rapid access to the actual world—to see only what we attend to and only attend to what we need—for the task at hand.

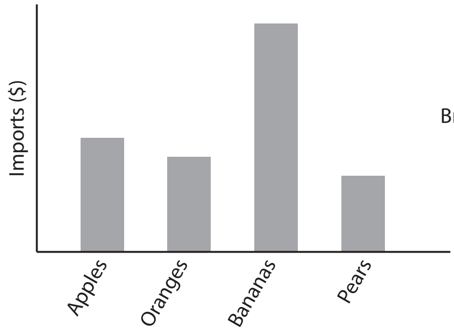
◆Kevin O'Regan's essay on the nature of the consciousness illusion brings into clear focus the fact that there is a major problem to be solved, how do we get a subjective impression of perceiving a detailed world, while all available evidence shows that we pick up very little information. It also points to the solution—just in time processing. J.K. O'Regan, 1992. Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*. 46: 461–488.

The one-tenth of a second or so that it takes to make an eye movement is such a short time in terms of the brain's neuron-based processing clock that it seems instantaneous. Our illusory impression that we are constantly aware of everything happens because our brains arrange for eye movements to occur and the particularly relevant information to be picked up just as we turn our attention to something we need. We do not have the whole visual world in conscious awareness. In truth, we have very little, but we can get anything we need through mechanisms that are rapid and unconscious. We are unaware that time has passed and cognitive effort has been expended. Exactly how we get the task-relevant information and construct meaning from it is a central focus of this book.

The understanding that we only sample the visual world on a kind of need-to-know basis leads to a profoundly different model of perception, one that has only emerged over the last decade or so as psychologists and neurophysiologists have devised new techniques to probe the brain.

According to this new view, visual thinking is a process that has the allocation of attention as its very essence. Attention, however, is multifaceted. Making an eye movement is an act of attending. The image on the retina is analyzed by further attention-related processes that tune our pattern-finding mechanisms to pull out the pattern most likely to help with whatever we are doing. At a cognitive level, we allocate scarce "working memory" resources to briefly retain in focal attention only to those pieces of information most likely to be useful. Seeing is all about attention. This new understanding leads to a revision of our thinking about the nature of visual consciousness. It is more accurate to say that we are conscious of the *field of information* to which we have *rapid access* rather than that we are immediately conscious of the world.

This new understanding also allows us to think about graphic design issues from a new and powerful perspective. We can now begin to develop a science of graphic design based on a scientific understanding of visual attention and pattern perception. To the extent to which it is possible to set out the message of this book in a single statement, the message is this: *Visual thinking consists of a series of acts of attention, driving eye movements and tuning our pattern-finding circuits.* These acts of attention are called *visual queries*, and understanding how visual queries work can make us better designers. When we interact with an information display, such as a map, diagram, chart, graph, or a poster on the wall, we are usually trying to solve some kind of cognitive problem. In the case of the map, it may be how to get from one location to another. In the case of the graph, it may be to determine the trend; for example, is the population increasing or decreasing over time? What is the shape of the trend? The answers to our questions can be obtained by a series of searches for particular patterns—visual queries.



To find out which kind of fruit import is the largest by dollar value we make visual queries to find the tallest bar, then find and read the label beneath.

To find a fast route, we first make visual queries to find the starting and ending cities, then we make queries to find a connected red line, indicative of fast roads, between those points.

At this point, you may be considering an obvious objection. What about the occasions when we are not intensely involved in some particular task? Surely we are not continually constructing visual queries when we are sitting in conversation with someone, or strolling along a sidewalk, or listening to music. There are two answers to this. The first is that, indeed, we are not always thinking visually with reference to the external environment; for example, we might be musing about the verbal content of a conversation we had over the telephone. The second is we are mostly unaware of just how structured and directed our seeing processes are. Even when we are in face-to-face conversation with someone, we constantly monitor facial expressions, the gestures and gaze direction of that person, to pick up cues that supplement verbal information. If we walk on a path along the sidewalk of a city, we constantly monitor for obstacles and choose a path to take into account the other pedestrians. Our eyes make anticipatory movements to bumps and stones that may trip us, and our brains detect anything that may be on a trajectory to cross our path, triggering an eye movement to monitor it. Seeing while walking is, except on the smoothest and most empty road, a highly structured process.

To flesh out this model of visual thinking, we need to introduce key elements of the apparatus of vision and how each element functions.

THE APPARATUS AND PROCESS OF SEEING

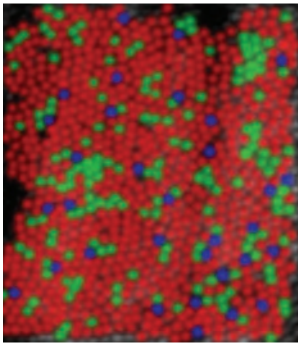
The eyes are something like little digital cameras. They contain lenses that focus an image on the eyeball. Many find the fact that the image is upside-down at the back of the eye to be a problem. But the brain is a computer, albeit quite unlike a digital silicon-based one, and it is as easy for the brain to compute with an upside-down image as a right-side-up image.

Just as a digital camera has an array of light-sensitive elements recording three different color values, so the eye also has an array of light-sensitive cones recording three different colors (leaving aside rods◆). The analogy goes still further. Just as digital cameras compress image data for more compact transmission and storage, so several layers of cells in the retina extract what is most useful. As a result, information can be transmitted from the 100 million receptors in the eye to the brain by means of only 1 million fibers in the optic nerve.

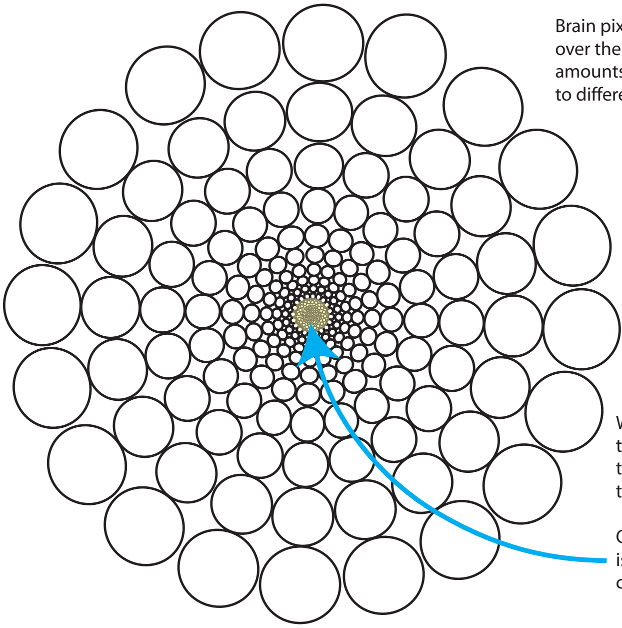
There is, however, a profound difference between the signal sent from the eye to the back of the brain for early-stage processing and the signal sent to a memory chip from the pixel array of a digital camera. *Brain pixels* are concentrated in a central region called the *fovea, whereas* camera pixels are arranged in a uniform grid. Also, brain pixels function as little image-processing computers, not just passive recorders.

Visual detail can only be seen via the fovea, at the very center of the visual field. Our vision is so good in this region that each eye can resolve about 100

◆The human eye actually contains four different receptor types, three cone types and rods. However, because rods function mainly only at low light levels, in our modern, brightly lit world we can for all practical purposes treat the eye as a three-receptor system. It is because of this that we need only three different wavelength receptors in digital cameras.

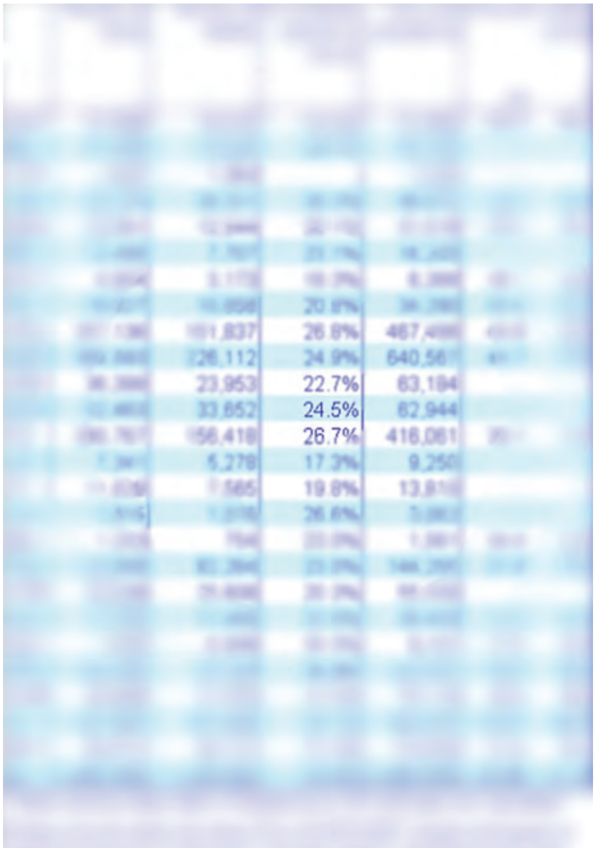


At the back of the eye is a mosaic of photo-receptors. Each responds to the amount of light falling on it. This example shows the central foveal region where three different cone receptors register different colors of light.

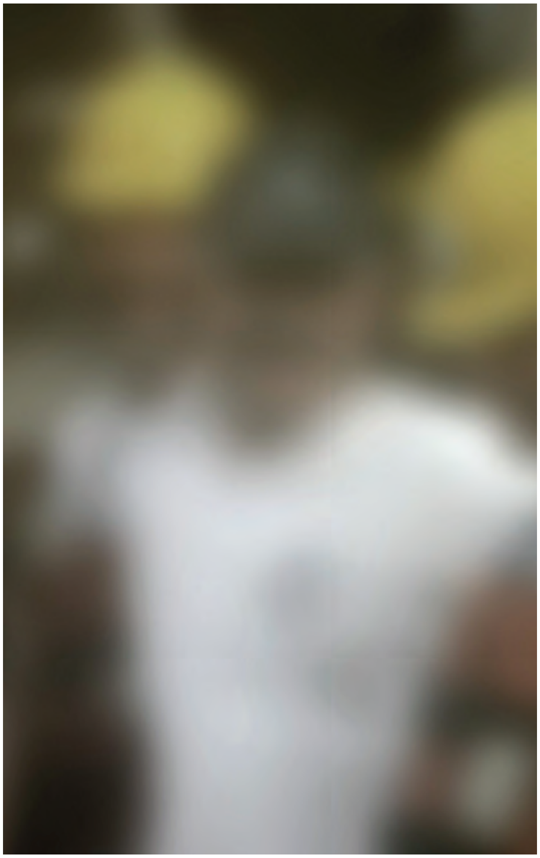


points on the head of a pin held at arm's length, but the region is only about the size of our thumbnail held at arm's length. At the edge of the visual field, vision is terrible; we can just resolve something the size of a human head. For example, we may be vaguely aware that someone is standing next to us, but unless we have already glanced in that direction we will not know who it is.

The non-uniformity of the visual processing power is such that half our visual brain power is directed to processing less than 5 percent of the visual world. This is why we have to move our eyes; it is the only way we can get all that brain power directed where it will be most useful. Non-uniformity is also one of the key pieces of evidence showing that we do not comprehend the world all at once. We cannot possibly grasp it all at once since our nervous systems only process details in a tiny location at any one instant.



We only process details in the center of the visual field. We pick up information by directing our foveas using rapid eye movements.

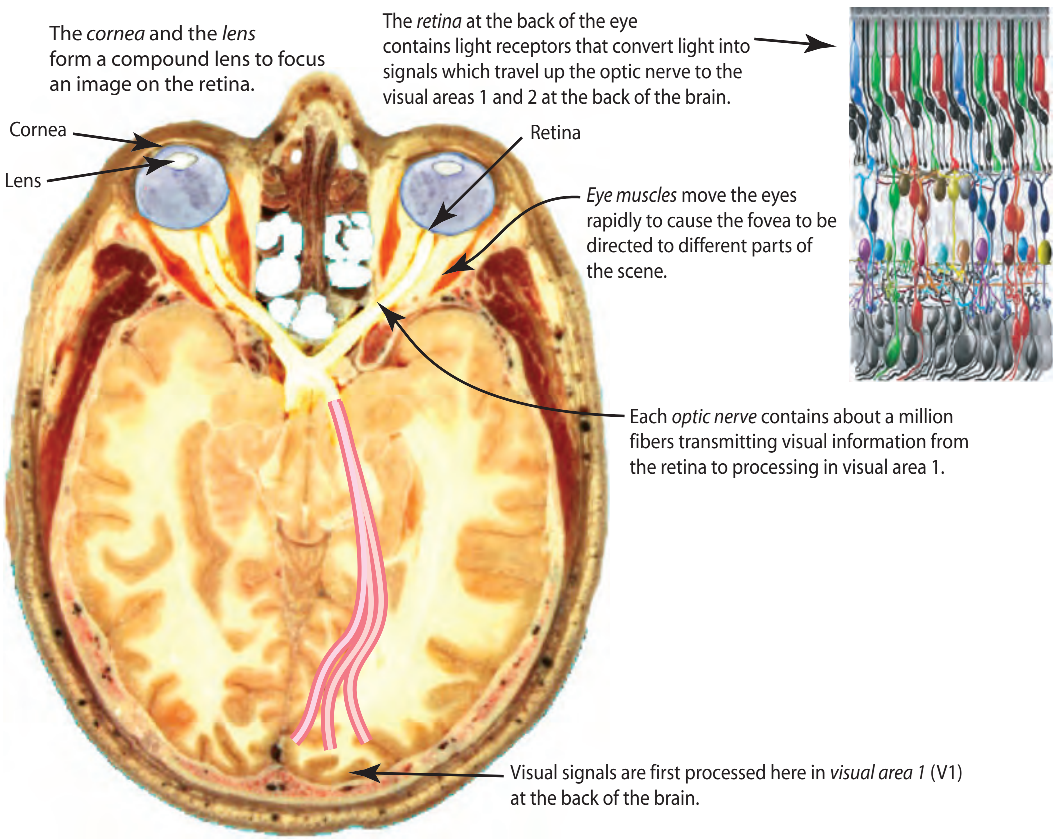


At the edge of the visual field, we can barely see that someone is standing next to us.

The term brain pixel was introduced earlier by way of contrast with digital camera pixels. Brain pixels provide a kind of distorted neural map covering the whole visual field. There are a great many tiny ones processing information from central regions where we direct our gaze and only a few very large ones processing information at the edge of the visual field. Because of this, we cannot see much out of the corners of our eyes.

Strong eye muscles attached to each eyeball rotate it rapidly so that different parts of the visual world become imaged on the central high-resolution fovea. The muscles accelerate the eyeball to an angular velocity up to 900 degrees per second, then stop it, all in less than one-tenth of a second. This movement is called a saccade, and during a saccadic eye movement, vision is suppressed. The eyes move in a series of jerks pointing the fovea at interesting and useful locations, pausing briefly at each, before flicking to the next point of interest. Controlling these eye movements is a key part of the skill of seeing.

We do not see the world as jerky, nor for the most part are we aware of moving our eyes, and this adds yet more evidence that we do not perceive what is directly available through our visual sense.



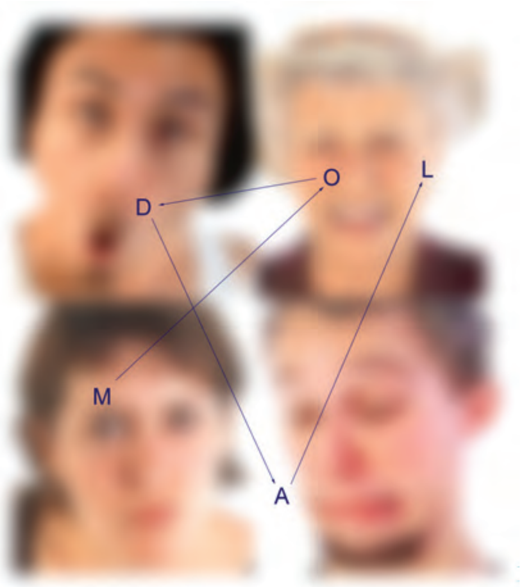
This illustration is based on a slice through the head of the visible man at the level of the eyeballs. Its color has been altered to show the eye muscles and the eyeballs more clearly.

THE ACT OF PERCEPTION

The visual field has a big hole in it. Cover your left eye and look at the X. Move the page nearer and farther away, being sure to keep the X and the B horizontally aligned.

X B

At some point the B should disappear. This is because the image of the B is falling on the blind spot, a region of the retina where there are no receptors at the point where the optic nerve and blood vessels enter the eye. We are unaware that we have this hole in our visual field. The brain does not know that it has a blind spot, just as it does not know how little of the world we see at each moment. This is more evidence that seeing is not at all the passive registration of information. Instead it is active and constructive.



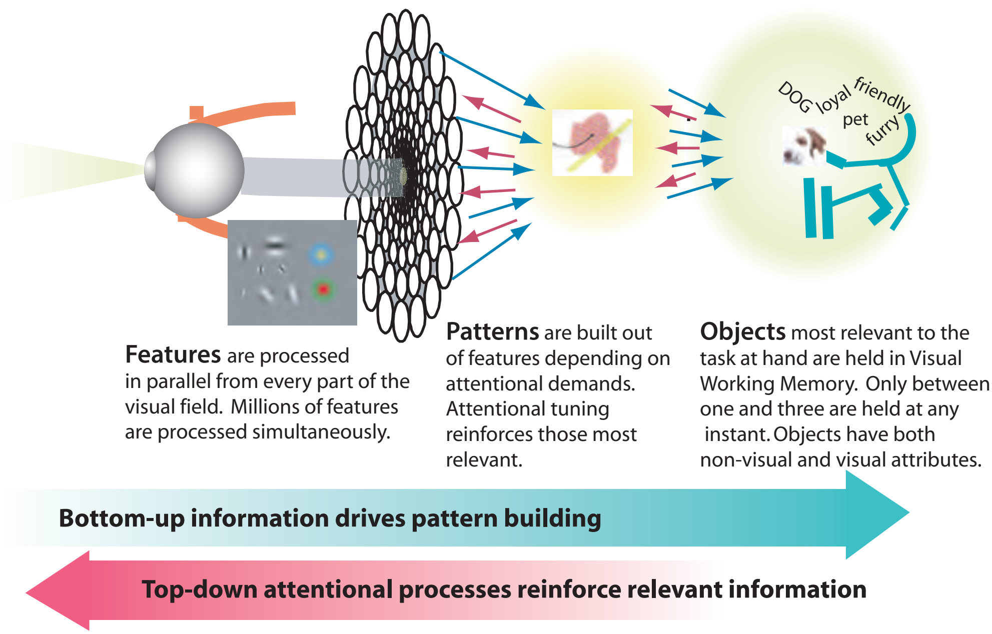
Broadly speaking, the act of perception is determined by two kinds of processes: *bottom-up*, driven by the visual information in the pattern of light falling on the retina, and *top-down*, driven by the demands of attention, which in turn are determined by the needs of the tasks. The picture shown above is designed to demonstrate how top-down attention can influence what you see and how.

First look at the letters and lines. Start with the M and follow the sequence of lines and letters to see what word is spelled. You will find yourself making a series of eye movements focusing your visual attention on the small area of each letter in turn. You will, of course, notice the faces in the background but as you perform the task they will recede from your consciousness.

Next look at the faces and try to interpret their expressions. You will find yourself focusing in turn on each of the faces and its specific features, such as the mouth or eyes, but also as you do this the letters and lines will recede from your consciousness. Thus, what you see depends on both the information in the pattern on the page as it is processed bottom-up through the various neural processing stages, and on the top-down effects of attention that determines both where you look and what you pull out from the patterns on the page.

There are actually two waves of neural activity that occur when our eye alights on a point of interest. An information-driven wave passes information first to the back of the brain along the optic nerve, then sweeps forward to the forebrain, and an attention-driven wave originates in the attention control centers of the forebrain and sweeps back, enhancing the most relevant information and suppressing less relevant information.

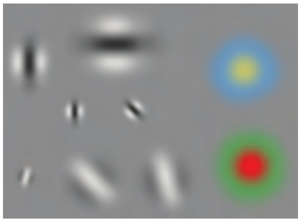
The neural machinery of the visual system is modular in the sense that distinct regions of the brain perform specific kinds of computations before passing the processed information on to some other region. The visual system has at least two dozen distinct processing modules, each performing some different computational task, but for the purposes of this overview we



will simplify to a three-stage model. The processing modules are organized in a hierarchy, with information being transferred both up and down from low-level brain-pixel processors to pattern and object processors. We shall consider it first from a bottom-up perspective, and then from a top-down perspective.

BOTTOM-UP

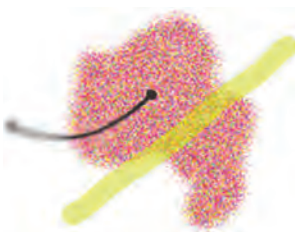
In the bottom-up view, information is successively selected and filtered so that meaningless low-level features in the first stage form into patterns in the second stage, and meaningful objects in the third stage.



Some neurons that process elementary features respond to little packets of orientation and size information. Others respond best to redness, yellowness, greenness, and blueness. Still others respond to different directions of motion.

The main *feature processing* stage occurs after information arrives in the V1 cortex, having traveled up the optic nerve. There are more neurons devoted to this stage than any other. Perhaps as many as five billion neurons form a massively parallel processing machine simultaneously operating on information from only one million fibers in the optic nerve. Feature detection is done by several different kinds of brain pixel processors that are arranged in a distorted map of visual space. Some pull out little bits of size and orientation information, so that every part of the visual field is simultaneously processed for the existence of oriented edges or contours. Others compute red-green differences and yellow-blue differences, and still others process the elements of motion and the elements of stereoscopic depth. The brain has sufficient neurons in this stage to process every part of the visual field simultaneously for each kind of feature information. In later chapters, we will discuss how understanding features processing can help us design symbols that stand out distinctly.

At the intermediate level of the visual processing hierarchy, feature information is used to construct increasingly complex *patterns*. Visual space is divided up into regions of common texture and color. Long chains of features become connected to form continuous contours. Understanding how this occurs is critical for design because this is the level at which space becomes organized and different elements become linked or segregated. Some of the design principles that emerge at this level have been understood for over seventy years through the work of Gestalt psychology (*gestalt* means form or configuration in German). But there is also much that we have learned in the intervening years through the advent of modern neuroscience that refines and deepens our understanding.



At the intermediate pattern finding stage of the visual system, patterns are formed out of elementary features. A string of features may form the boundary of a region having a particular color. The result is that the visual field is segmented into patterns.

At the top level of the hierarchy, information that has been processed from millions upon millions of simple features has been reduced and distilled through the pattern-processing stage to a small number of visual *objects*. The system that holds about three objects in attention at one time is called *visual working memory.* The small capacity of visual working memory is the reason why, in the experiment described at the start of this chapter, people failed to recognize that they were speaking to a different person. The information about the person they were talking to became displaced from their visual (and verbal) working memories by more immediate task-relevant data.

Although something labeled "dog" might be one of the objects we hold in our visual working memory, there is nothing like a picture of a dog in the head; rather we have a few visual details of the dog that have been recently fixated. These visual details are linked to various kinds of information that we already know about dogs through a network of association, and therein lies the power of the system. Concepts that dogs are loyal, pets, furry, and friendly may become activated and ready for use. In addition, various possibilities for action may become activated, leading to a heightened state of readiness. Actions such as petting the dog or avoiding the dog (depending on our concepts) become primed for activation. Of course if it is our own dog, "Millie," a much richer set of associations become activated and the possibilities for action more varied. This momentary binding together of visual information with nonvisual concepts and action priming is central to what it means to perceive something.

When we see something, such as a dog, we do not simply form an image of that dog in our heads. Instead, the few features that we have directly fixated are bound together with the knowledge we have about dogs in general and this particular dog. Possible behaviors of the dog and actions we might take in response to it are also activated.

The reason why we can make do with only three or four objects extracted from the blooming buzzing confusion of the world is that these few objects are made up of exactly what we need to help us perform the task of the moment. Each is a temporary nexus of meaning and action. Sometimes nexus objects are held in mind for a second or two; sometimes they only last for a tenth of a second. The greatly limited capacity of visual working memory is a major bottleneck in cognition, and it is the reason why we must often rely on external visual aids in the process of visual thinking.

It is tempting to think of visual working memory as the place real visual thinking occurs, but this is a mistake. One reason it is easy to think this way is that this is the way computers work. In a digital computer, all complex operations on data occur in the central processing unit. Everything else is about loading data, getting it lined up so that it is ready to be processed just when it is needed, and sending it back out again. The brain is not like this. There is actually far more processing going on in the lower-level feature processing and pattern-finding systems of the brain than in the visual working memory. It is much more accurate to think of visual thinking as a multicomponent cognitive system. Each part does something that is relatively simple. For example, the intermediate pattern processors detect and pass on information about a particular red patch of color that happens to be imaged on a particular part of the retina. An instant later this red patch may come to be labeled as "poppies."

In many ways, the real power of *visual* thinking rests in pattern finding. Often to see a pattern is to find a solution to a problem. Seeing the path to the door tells us how to get out of the room, and that path is essentially a kind of visual pattern. Similarly, seeing the relative sizes of segments in a pie chart tells us which company has the greatest market shares.

Responses to visual patterns can be thought of as another type of pattern. (To make this point we briefly extend the use of the word "pattern" beyond its restricted sense as something done at a middle stage in visual processing.) For most mundane tasks we do not think through our actions from first principles. Instead, a response pattern like walking towards the door is triggered from a desire to leave the room. Indeed it is possible to think of intelligence in general as a collaboration of pattern-finding processors.◆ A way of responding to a pattern is also a pattern, and usually one we have executed many times before. A very common pattern of seeing and responding is the movement of a mouse cursor to the corner of a computer interface window, together with a mouse click to close that window. Response patterns are the essence of the skills that bind perception to action. But they have their negative side, too. They also cause us to ignore the great majority of the information that is available in the world so that we often miss things that are important.

◆This view of intelligence as a kind of hierarchy of pattern finding systems has been elaborated in *On Intelligence* by Jeff Hawkins and Sandra Blakeslee (Times Books, New York, 2004).

TOP-DOWN

So far we have been focusing on vision as a bottom-up process:

$$\text{retinal image} \rightarrow \text{features} \rightarrow \text{patterns} \rightarrow \text{objects}$$

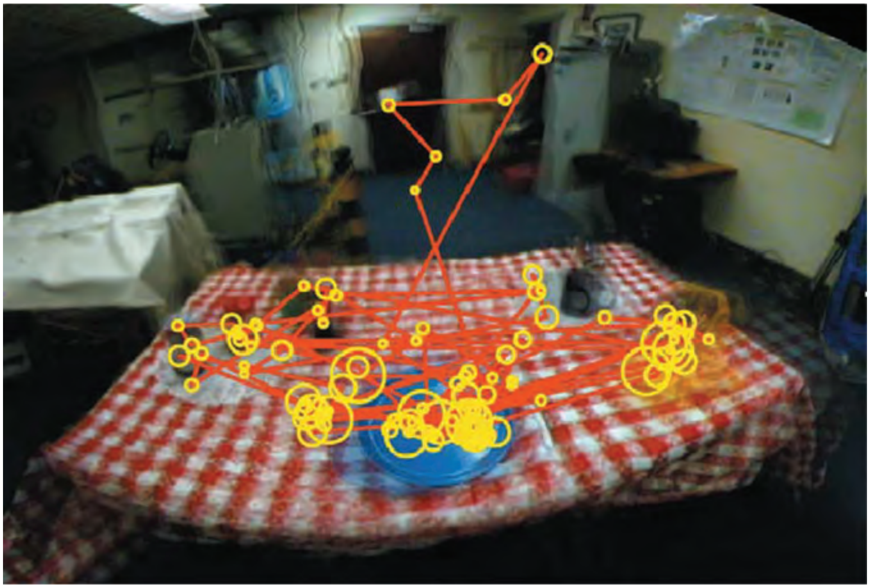
but every stage in this sequence contains corresponding top-down processes. In fact, there are more neurons sending signals back down the hierarchy than sending signals up the hierarchy.

We use the word *attention* to describe top-down processes. Top-down processes are driven by the need to accomplish some goal. This might be an action, such as reaching out and grasping a teacup or exiting a room. It might be a cognitive goal, such as understanding an idea expressed in a diagram. There is a constant linking and re-linking of different visual information with different kinds of nonvisual information. There is also a constant *priming* of action plans (so that if we have to act, we are ready) and action plans that are being executed. This linking and re-linking is the essence of high-level attention, but it also has implications for other lower-level processes.

At the low level of feature and elementary pattern analysis, top-down attention causes a bias in favor of the signals we are looking for. If we are looking for red spots then the red spot detectors will signal louder. If we are looking for slanted lines then slanted line feature detectors will have their signal enhanced. This biasing in favor of what we are seeking or anticipating occurs at every stage of processing. What we end up actually perceiving is the result of information about the world strongly biased according to what we are attempting to accomplish.

Perhaps the most important attentional process is the sequencing of eye movements. Psychologist Mary Hayhoe and computer scientist Dana Ballard collaborated in using a new technology that tracked individuals' eye movements while they were able to move freely.◆ This allowed them to study natural eye movements "in the wild" instead of the traditional laboratory setup with their heads rigidly fixed in a special apparatus.

◆Mary Hayhoe and Dana Ballard. 2005. Eye movements in natural behavior. *Trends in Cognitive Science*. 9(4): 188–194.



The sequence of eye movements made by someone making a peanut butter and jelly sandwich. The yellow circles show the eye fixations. (Image courtesy of Mary M. Hayhoe.)

They had people carry out everyday tasks, such as making a peanut butter and jelly sandwich, and discovered a variety of eye movement patterns. Typically, people exhibited bursts of rapid eye movements when they first encountered the tools and ingredients laid out in front of them. This presumably allowed them to get a feel for the overall layout of the workspace. Each of these initial fixations was brief, usually one-tenth of a second or less. Once people got to work, they would make much longer fixations so that they could, for example, spread the peanut butter on the bread. Generally, there was great economy in that objects were rarely looked at unnecessarily; instead, they were fixated using a "just-in-time strategy." When people were performing some action, such as placing a lid on a jar, they did not look at what their hands were doing but looked ahead to the jar lid while one hand moved to grasp it. Once the lid was in hand, they looked ahead to fixate the top of the jar enabling the next movement of the lid. There were occasional longer-term look-aheads, where people would glance at something they might need to use sometime in the next minute or two. The overall impression we get from this research is of a remarkably efficient, skilled visual process with perception and action closely linked—the dominant principle being that we only get the information we need, when we need it.

How do we decide where to move our eyes in a visual search task? If our brains have not processed the scene, how do we know where to look? But if we already know what is there, why do we need to look? It's a classic chicken and egg problem. The system seems to work roughly as follows.◆ Part of our brain constructs a crude map of the characteristics of the information that we need in terms of low-level features. Suppose I enter a supermarket produce section looking for oranges. My brain will tune my low-level feature receptors so that orange things send a stronger signal than patches of other colors. From this, a rough map of potential areas where there may be oranges will be constructed. Another part of my brain will construct a series of eye movements to all the potential areas on this spatial map. The eye movement sequence will be executed with a pattern processor checking off those areas where the target happened to be mangoes, or something else, so that they are not visited again. This process goes on until either oranges are found, or we decide they are probably hidden from view. This process, although efficient, is not always successful. For one thing, we have little color vision at the edges of the visual field, so it is necessary to land an eye movement near to oranges for the orange color-tuning process to work. When we are looking for bananas a shape-tuning process may also come into play so that regions with the distinctive curves of banana bunches can be used to aid the visual search.●

◆J.R. Duhamel, C.L. Colby, and M.E. Goldberg, 1992. The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*. Jan. 3; 255(5040):90–92.

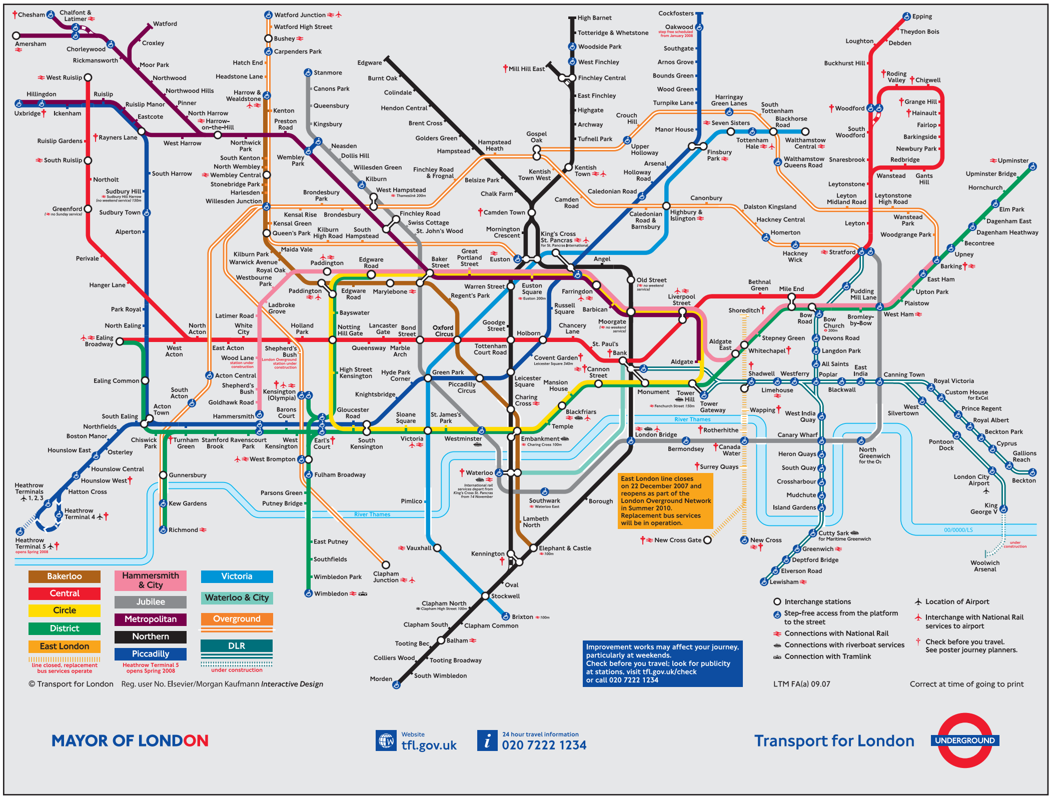
●The book *Active Vision* by John Findlay and Ian Gilchrist (Oxford University Press, 2003) is an excellent introduction to the way eye movements are sequenced to achieve perception for action.

IMPLICATIONS FOR DESIGN

If we understand the world through just-in-time visual queries, the goal of information design must be to design displays so *that visual queries are processed both rapidly and correctly for every important cognitive task the display is intended to support.* This has a number of important ramifications for graphic design. The first is that in order to do successful design we must understand the cognitive tasks and visual queries a graphic is intended to support. This is normally done somewhat intuitively, but it can also be made explicit.

A map is perhaps the best example illustrating how graphic design can support a specific set of visual queries. Suppose we are lodged in a hotel near Ealing in West London, and we wish to go to a pub near Clapham Common where we will meet a friend. We would do well to consider the underground train system and this will result is our formulating a number of cognitive tasks. We might like to know the following:

- Which combination of lines will get us to the pub?
- If there is more than one potential route, which is the shortest?
- What are the names of stations where train changes are needed?
- How long will the trip take?
- What is the distance between the hotel and its nearest underground station?
- What is the distance between the pub and its nearest underground station?
- How much will it cost?
- Many of these tasks can be carried out through visual thinking with a map.



Reproduced with kind permission of © Transport for London.

The famous London Underground map designed by Harry Beck is an excellent visual tool for carrying out some, but not all, of these tasks. Its clear schematic layout makes the reading of station names easy. The color coding of lines and the use of circles showing connection points make it relatively easy to visually determine routes that minimize the number of stops. The version shown here also provides rapidly accessible information about the fare structure through the grey and white zone bars.

Since we have previously used maps for route planning, we already have a cognitive plan for solving this kind of problem. It will likely consist of a set of steps something like the following.

STEP ONE is to construct a visual query to locate the station nearest our hotel. Assuming we have the name of a station, this may take quite a protracted visual search since there are more than two hundred stations on the map; if, as is likely, we already have a rough idea where in London our hotel is located, this will narrow the search space.

STEP TWO is to visually locate a station near the pub, and this particular task is also not well supported. The famous map does support the station-finding task by spacing the labels for clear reading, but unlike other maps it does not give an index or a spatial reference grid.

STEP THREE is to find the route connecting our start station and our destination station. This visual query is very well supported by the map. The lines are carefully laid out in a way that radically distorts geographical space, but this is done to maintain clarity so that the visual tracing of lines is easy. Color coding also supports visual tracing, as well as providing labels that can be matched to the table of lines at the side.

Suppose that the start station is Ealing on the District Line (green) and the destination station is Clapham Common on the Northern Line (black). Our brains will break this down into a set of steps executed roughly as follows. Having identified the Ealing station and registered that it is on the green-colored District Line, we make a series of eye movements to trace the path of this green contour. As we do so, our top-down attentional mechanisms will increase the amplification on neurons tuned to green so that they "shout louder" than those tuned to other colors, making it easier for our mid-level pattern finder to find and connect parts of the green contour. It may take several fixations to build the contour, and the process will take about two seconds. At this point, we repeat the tracing operation starting with Clapham Common, our destination station, which is on the black Northern Line. As we carry out this tracing, a second process may be operating in parallel to look for the crossing point with the green line. These operations might take another second or two.

Because of the very limited capacity of visual working memory, most information about the green contour (District Line) will be lost as we trace out the black contour (Northern Line). Not all information about the green line is lost, however; although its path will not have been retained explicitly, considerable savings occur when we repeat an operation such as this. Repeating a tracing operation will take less cognitive effort, and require fewer fixations, than finding it in the first place. A hallmark of visual thinking is that it is often easier to redo some cognitive operation than to remember it.

Of course, the process may not be quite as straightforward as described. Much can go wrong. We may be visually sidetracked by the wrong branch of the Northern Line. The process is flexible and adaptive, and good visual thinkers have error-checking procedures. A final scan of the entire route will confirm that it is, indeed, a valid solution.

STEP FOUR is a visual query to get a rough estimate of how many stations there are on the route. This is unlikely to involve actual counting; instead, it will be a judgment that will be used together with prior experience to produce an estimate of the journey duration. This will naturally lead to misjudgments because the lengths of lines on the map do not correspond to travel times.

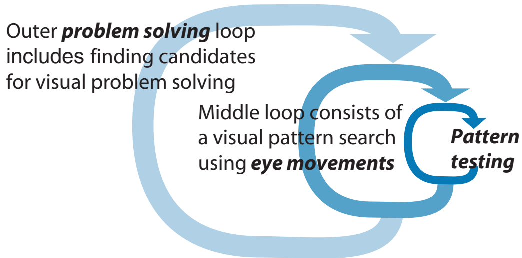
Our analysis shows that the London Underground map supports the visual route-finding step well but almost completely fails to support other planning steps such as finding a tube station near to the pub and estimating journey time. This is not to say that it is a bad design; the map is justifiably famous. The designers sacrificed spatial accuracy in favor of clear labels and routes, and therefore the map is very poor for providing information about the distance covered on the ground.

Having a computer behind an *interactive* graphic display takes the capability of visual cognitive tools to another level. A program called tubeguru™ is an Internet-based computer program that adds many of the cognitive supports that the underground map lacks; a local street map in the vicinity of a selected station helps with the start and end of journeys, and a journey planner provides information about how to get from a specific address to another specific address, using London's busses and trains.

Tubeguru and other computer programs with visual interfaces add an additional dimension to the cognitive process. They allow for some parts of the computation process to occur in the brain, and other parts to occur in the computer program. Human and computer together form a *problem-solving system* with the screen display and the keyboard providing the interface. In later chapters, we further explore this rather mind-bending idea. For now we return to the simpler case of visual thinking with a static graphic display.

NESTED LOOPS

A useful way of describing the way the brain operates to solve problems is as a set of nested loops. Outer loops deal with generalities. Inner loops process the details. In the outer loop, the brain constructs a set of steps to solve the problem and then executes them: find a map, find end point stations, trace lines, find intersections. This sequencing of the problem components is not in and of itself visual in nature; it is more likely to occur through the operations of verbal language subsystems of the brain. However, some problem components may be identified as having visual-thinking solutions, and these are used to construct visual queries on the display.



The middle loop on the diagram is a visual search executed to find patterns that address the visual query. This involves executing a sequence of eye movements. In the previous example, this involved a visual tracing of the colored contours that represent train lines.

The inner loop is activated when the eye arrives at a point of fixation. A process of visual testing begins, and patterns within the central region of the visual field are evaluated at a rate of about twenty per second; although since the eye only stays in one place for less than two-tenths of a second, roughly one to four simple patterns may be evaluated on each fixation. In the example of finding the green line, this evaluation might have the following form. Is this contour green? Is it a continuation of the section of contour that was just registered?

Of course we are being metaphoric here. Nothing in the brain is making up questions and answering them. Rather, a neural process produces a signal that results in a change in the status of objects held in visual working memory. One such change in status would be that a neural representation of a green contour object becomes "bound" to another section of green contour in the temporary nexus and held together through the process called visual working memory.

Nested processes are characteristic of computer programs, but the ones executed by the brain are far more flexible and adaptive than those executed in a computer. They rely on patterns of action that have been built up over experience. We already have encoded in our memories visual search patterns for a huge number of situations, such as what we need to look for when entering a restaurant, looking at a web page, and driving a car toward an intersection. These are not, however, rigid rules, but rather like flexible plans that can be adapted to particular circumstances.

DISTRIBUTED COGNITION

This chapter has alluded to the existence of various brain structures and processes. In later chapters, specific areas of the brain where processes occur will be discussed in more detail. We will encounter visual areas, V1 and V2 specialized for parallel feature processing, V4 specialized for pattern processing, the Fusiform Gyrus specialized for object processing, and the frontal lobes as well as midbrain areas specialized for controlling high-level attention. The control of eye movements is mediated by mechanisms in the lateral interparietal area and superior colliculus. There are several dozen areas of the brain that have been mapped in animals using recordings from single neurons. Recent advances in brain scanning have allowed researchers to see which areas of the brain are most active when people perform specific visual tasks. The latest techniques even reveal the sequence in which different areas become active.

This is not, however, a book about neurophysiology, but a book about visual thinking. Its purpose is to provide a theoretical understanding of how we perceive in order to inform the design process. The names are useful but ultimately it does not matter exactly where in the brain something happens; instead it is important to know what kinds of visual information the brain can process efficiently. The timing and sequencing of visual operations is also important. Therefore, the approach taken in this book is functional and modular. Specific brain structures are mentioned because each has a specialized function in the visual thinking process. We could just talk about these brain structures as abstract functioning modules, but giving them their proper names preserves the link with the underlying science. Overall the goal is to steer a path between oversimplification and overwhelming neurophysiological detail.

There is a general point to be considered regarding the modular structure of the brain. The brain is not a holistic undifferentiated processor of information, as was once thought. Instead, it is made up of a number of specialized regions, each devoted to carrying out some specific processing task, such as scheduling eye movements or processing features. Visual thinking comes about through the coordinated activities of these processing regions. There is no central processing unit in the brain; rather the whole functions as a kind of *distributed* computer. Also, there is no central conductor coordinating all of the distributed parts. Each component coordinates interactions with the other components to which it is connected. This is the reason why we can sometimes carry out more than one task simultaneously—talking and driving a car, for example. These tasks use regions of the brain that are mostly independent.

The idea of distributed cognition is no longer particularly controversial, but the idea of distributed cognition begs to be extended outside the head, and this is truly radical. Distributed cognition holds that cognition is the result of a set of interconnected processing modules, each doing something relatively simple and sending signals on to other modules. But why must these modules be inside the head? What is the difference between information stored in human memory and information stored in a book or picture? After all, we can store images in a photo album far better than we can store them in our head. We can store them even better on a computer if it has a effective search capability. Thinking involves a constant interplay between new patterns and old patterns, and patterns can come from both inside and outside the skull. As Don Norman famously noted,◆ *"The power of the unaided mind is highly overrated. Without external aids, memory, thought and reasoning are all constrained. But human intelligence is highly flexible and adaptive, superb at inventing procedures and objects that overcome its own limits. The real powers come from devising external aids that enhance cognitive activities. How have we increased memory, thought and reasoning? By the invention of external aids: it is things that make us smart."*

◆Donald A. Norman, 1993. Cognition in the head and in the world. *Cognitive Science*. 17: 1–6.

Graphs, diagrams, and illustrations have only become widely available as visual thinking tools over the past two hundred years. More recently there has been an explosive development of diagramming techniques driven first by color printing technology and currently by the Internet. Graphic PowerPoint slides and the like have become a ubiquitous tool for information presentation. Increasingly, the tools that support cognition are computer-based, and increasingly they incorporate images and visualizations as well as words. The term *visualization* as it is used in the previous sentence is actually quite new. Visualizations used to be mental images that people formed while they thought. Now the term more often means a graphical representation of some data or concepts. Visualizations are becoming important in most areas of science and commerce. Advertisements have long been visualizations. All of these artifacts are tools for visual thinking.

CONCLUSION

Perceiving is a skilled active process. We seek out what we need through frequent eye movements, so that critical information falls on the high-resolution fovea. Eye movements are executed to satisfy our need for information and can be thought of as a sequence of visual queries on the visual world. Each time the eye briefly comes to rest, the pattern-processing machinery goes to work sorting out what is most likely to be relevant to our current cognitive task. Almost everything else is either not seen at all or retained for only a fraction of a second. A few fragments are held for a second or two, and a tiny percentage forms part of our long-term memory. To be sure we do have visual memories built up over our lifetimes and these are what make the scraps of information we capture meaningful and useful. These memories are not detailed, but they do provide frameworks for fleshing out the fragments. They also provide frameworks for planning eye movements and other kinds of actions.

The idea of the visual query is shorthand for what we do when we obtain information either from the world at large or some kind of information display. We make visual queries every time we search for some visual information that we need to carry out our cognitive task of the moment. Understanding what visual queries are easily executed is a critical skill for the designer. The special skill of designers is not so much skill with drawing or graphic design software, although these are undoubtedly useful, but the talent to analyze a design in terms of its ability to support the visual queries of others. This talent comes from hard-won pattern analysis skills that become incorporated into the neural fabric of perception, as well as the skill to execute a cognitive process that takes into account a variety of competing considerations.

One reason design is difficult is that the designer already has the knowledge expressed in the design, has seen it develop from inception, and therefore cannot see it with fresh eyes. The solution is to be analytic and this is where this book is intended to have value. Effective design should start with a visual task analysis, determine the set of visual queries to be supported by a design, and then use color, form, and space to efficiently serve those queries. Skilled graphic designers already do this intuitively. It is my earnest hope that this book will help in the transition from unskilled to skilled designer by providing visual analytic tools derived from an up-to-date understanding of human perception.

At this stage, it should be clear that this book is not about the kind of visual thinking that goes into fine art where the goals are frequently the opposite of clarity, but rather beauty, visual impact, or an investigation of a new vocabulary of expression. Because of its exploratory, pioneering nature, leading-edge art often speaks only to small cliques of insiders, collectors, critics, and gallery owners. This book is about graphic design that provides a channel for clear communication that supports visual thinking and acts as an interface to the vast information resources of the modern world. This should not, however, imply that its message is the enemy of creative expression. There is never a single optimal solution to a design problem, but rather a huge variety of alternative clear and effective designs.