VISUAL THINKING *for* DESIGN

Colin Ware

Preface

There has been a revolution in our understanding of human perception that goes under the name "active vision." Active vision means that we should think about graphic designs as cognitive tools, enhancing and extending our brains. Although we can, to some extent, form mental images in our heads, we do much better when those images are out in the world, on paper or computer screen. Diagrams, maps, web pages, information graphics, visual instructions, and technical illustrations all help us to solve problems through a process of visual thinking. We are all cognitive cyborgs in this Internet age in the sense that we rely heavily on cognitive tools to amplify our mental abilities. Visual thinking tools are especially important because they harness the visual pattern finding part of the brain. Almost half the brain is devoted to the visual sense and the visual brain is exquisitely capable of interpreting graphical patterns, even scribbles, in many different ways. Often, to see a pattern is to understand the solution to a problem.

The active vision revolution is all about understanding perception as a dynamic process. Scientists used to think that we had rich images of the world in our heads built up from the information coming in through the eyes. Now we know that we only have the illusion of seeing the world in detail. In fact the brain grabs just those fragments that are needed to execute the current mental activity. The brain directs the eyes to move, tunes up parts of itself to receive the expected input, and extracts exactly what is needed for our current thinking activity, whether that is reading a map, making a peanut butter and jelly sandwich, or looking at a poster. Our impression of a rich detailed world comes from the fact that we have the capability to extract anything we want at any moment through a movement of the eye that is literally faster than thought. This is automatic and so quick that we are unaware of doing it, giving us the illusion that we see stable detailed reality everywhere. The process of visual thinking is a kind of dance with the environment with some information stored internally and some externally and it is by understanding this dance that we can understand how graphic designs gain their meaning.

Active vision has profound implications for design and this is the subject of this book.

It is a book about how we think visually and what that understanding can tell us about how to design visual images. Understanding active vision tells us which colors and shapes will stand out clearly, how to organize space, and when we should use images instead of words to convey an idea.

Early on in the writing and image creation process I decided to "eat my own dog food" and apply active vision-based principles to the design of this book. One of these principles being that when text and images are related they should be placed in close proximity. This is not as easy as it sounds. It turns out that there is a reason why there are labeled figure legends in academic publishing (e.g. Figure 1, Figure 2, etc.). It makes the job of the compositor much easier. A compositor is a person whose specialty is to pack images and words on the page *without reading the text*. This leads to the labeled figure and the parenthetical phrase often found in academic publishing, "see Figure X". This formula means that Figure X need not be on the same page as the accompanying text. It is a bad idea from the design perspective and a good idea from the perspective of the publisher. I decided to integrate text and words and avoid the use of "see Figure X" and the result was a difficult process and some conflict with a modern publishing house that does not, for example, invite authors to design meetings, even when the book is about design. The result is something of a design compromise but I am grateful to the individuals at Elsevier who helped me with what has been a challenging exercise.

There are many people who have helped. Diane Cerra with Elsevier was patient with the difficult demands I made and full of helpful advice when I needed it. Denise Penrose guided me through the later stages and came up with the compromise solution that is realized in these pages. Dennis Schaefer and Alisa Andreola helped with the design. Mary James and Paul Gottehrer provided cheerful and efficient support through the detailed

editing the production process. My wife, Dianne Ramey, read the whole thing twice and fixed a very great number of grammar and punctuation errors. I am very grateful to Paul Catanese of the New Media Department at San Francisco State University and David Laidlaw of the Computer Graphics Group at Brown University who provided content reviews and told me what was clear and what was not. I did major revisions to Chapters 3 and 8 as a result of their input.

This book is an introduction to what the burgeoning science of perception can tell us about visual design. It is intended for anyone who does design in a visual medium and it should be of special interest to anyone who does graphic design for the internet or who designs information graphics of one sort or another. Design can take ideas from anywhere, from art and culture as well as particular design genres. Science can enrich the mix.

Colin Ware

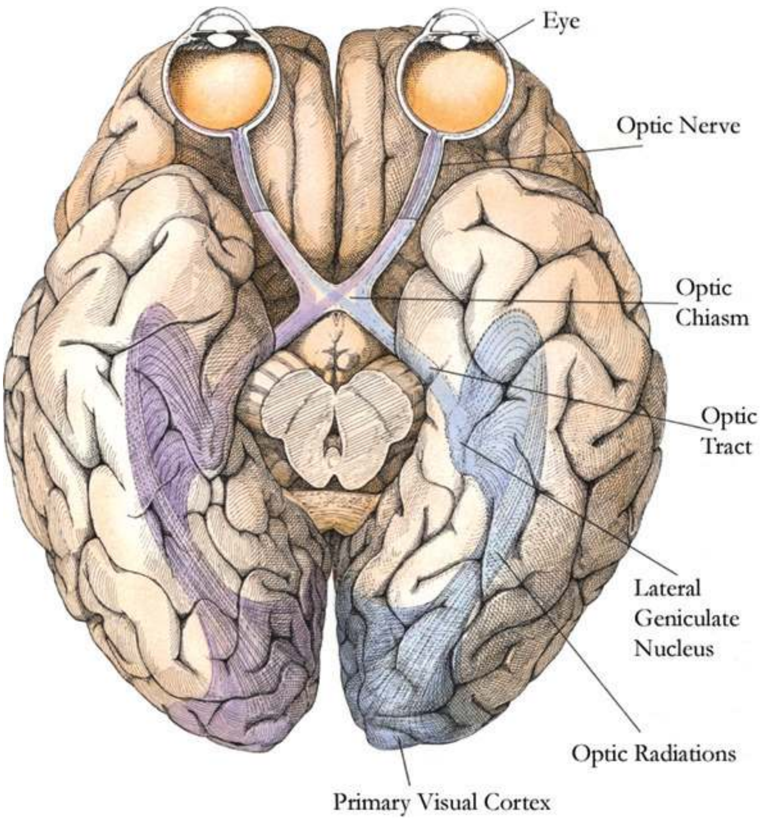
January 2008

Chapter 2

What We Can Easily See

Imagine a thief searching a moonlit room with a flashlight. He can see the vague outlines of furniture, but no details at all. He uses the narrow beam of the flashlight to pick out the tops of dressers and side tables where he expects valuables may be. Our visual attention actually works something like this; we move the spotlight of our attention by moving our eyes from point to point, picking out details. Generally we only have vague information to plan each eye movement, so we often fail to find the information we seek on a particular eye fixation. But some things are very easy to find even at the edge of the visual field, like a blinking light over the water, or a bright patch of red sweater in a crowd of people wearing black. We can sense them in the periphery and then make an eye movement so that they become the center of both vision and attention.

This chapter is about the theory of vision that describes what makes something small easy to see. It is also about the nitty-gritty details of design. What does it take to make a graphic symbol that can be found rapidly? How can something be highlighted? The problem for the designer is to ensure all visual queries can be effectively and rapidly served. In practice, this means ensuring that semantically meaningful graphic objects that make up a design each has the right amount of salience. The most important and frequent visual queries should be supported with the most visually distinct objects. The perceptual laws of visual distinctness are based on the low-level early-stage processing in the visual system. The elementary pattern-processing systems we find there provide the substrate on which all graphical interpretation is built.



A view of the brain from beneath. Light is transformed into neural signals by the retina. Information is then passed along the optic nerve, via the waystation of the lateral geniculate nucleus to the primary visual cortex located at the back of the brain.

In the primary visual cortex billions of neurons process the entire image, providing the elements of form, color, texture, stereoscopic depth, and motion all at once.

In the following set of random letters, two p's have been highlighted with a blue background. Finding those two p's is easy. They seem to pop out from the page. In contrast, it is more difficult to find the two q's. To find the q's every letter must be scanned and it will take at least ten times longer than finding the p's. Further, finding the q's will impose a much greater cognitive burden, certainly disrupting one's train of thought.

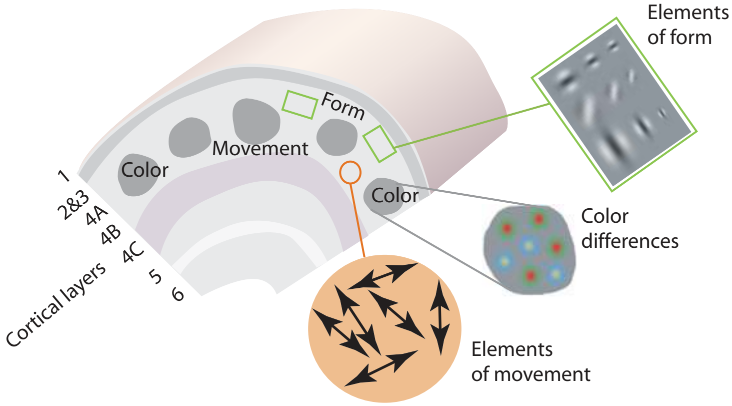
ehklhfdiyaioryweklblkhockxlyhirhupwerlkhlkuyxoiasusifdh
lksajdhflkihqdaklljerlajesljselusdslfjsalsuslcjlsdsjaf;ljdulafjluj
oufojrtopjhklghqlkshlkfhlkdshflymcvciwopzlsifhrmckreieui

The fact that the p's are easy to see seems straightforward and unsurprising. But why do the q's take longer to find? After all, we all spend thousands of hours reading and writing with this alphabet. So the q shape is not unfamiliar.

THE MACHINERY OF LOW-LEVEL FEATURE ANALYSIS

In order to understand why the p's are easy and the q's are hard on the previous page we need to dig more deeply into what is going on during the earlier stages of visual processing. The neural architecture of the primary visual cortex has been mapped in detail through experiments in which neurophysiologists inserted very small electrical probes into single neurons in the brains of various animals. David Hubel and Torsten Weisel pioneered this method in the 1960s. They discovered that at the back of the brain in a region called the primary visual cortex (also called visual area 1, or V1) cells would "fire," thereby emitting a series of spikes of electrical current when certain kinds of patterns were put in front of an animal's eyes.◆ Through many hundreds of such experiments, neuroscientists have discovered that V1 is a kind of tapestry of interlocking regions where different kinds of information are processed. The elements of color, shape, texture, motion, and stereoscopic depth are processed by different interwoven areas. Visual area 2 (V2) receives input from V1. V2's neurons respond to slightly more complex patterns, based on the processing already done in V1.

◆David Hubel and Torsten Wiesel obtained a Nobel prize for this groundbreaking research in 1981.



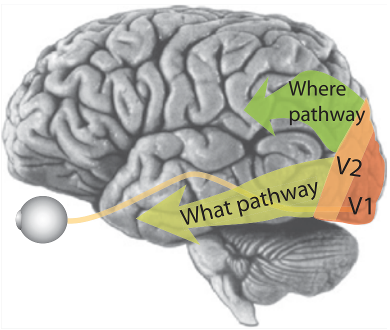
A fold of the primary visual cortex (V1) contains many layers of neurons. Signals from the retina, via the lateral geniculate nucleus, come up through several layers, dense with cross connections to layers 3 and 4 where individual neurons each respond to a specific simple feature.

There are a million nerve fibers providing input from each eye to the V1 area where several billion neurons process information. Information is then passed on to area V2 where several billion more neurons process it at a more complex level. Areas V1 and V2 can each be thought of as a parallel computer, far more complex and powerful than anything humans have built to date. Parallel computing is the name given to computer systems that are made up of many small computers, each doing essentially the same thing on a different piece of data. These kinds of computers are used to solve problems such as weather forecasting, where each processor deals with a little bit of the atmosphere. These cortical areas are *parallel* computers because they process every part of the visual image simultaneously, computing local orientation information, local color difference information, local size information, and local motion information.

WHAT AND WHERE PATHWAYS

In this chapter, we are mostly concerned with what goes on in the primary visual cortex, but it is useful to look ahead a bit to see where the information is going. V1 and V2 provide the inputs to two distinct processing systems called the *what* and the *where* systems, respectively.◆

◆Area V3 is also part of both what and where pathways but its processing role is still uncertain.



The what pathway sweeps forward from V1 and V2 along the lower edge of the brain on each side, processing information about the identity of an object.

The where pathway sweeps forward higher up on the brain and processes information about where objects in the world are located.

The what pathway is concerned with the identification of objects in the environment. It helps identify if a particular pattern of light and color represents a chicken, a Volkswagen, or Aunt Mabel. The where pathway is concerned with the location of information and with guiding actions in the world, such as reaching out and grabbing things, moving from place to place, or making eye movements. What makes an object easy to find is how easily we can direct a rapid eye movement to focus our attention on it.

EYE MOVEMENT PLANNING

Visual search is not random; however, there is a chicken and egg problem. If we are looking for something smallish, we can only see it when we are looking at it. But how do the eyes get directed to the right locations if the information has not been processed? The answer is that there is very limited pre-processing that is used to direct attention. To understand this is to understand what is easy to see.●

●The theory of visual search presented here is from J.M. Wolf, 1998. Visual search. In *Attention*. H. Paschler, Ed. Psychology Press.

Part of the machinery is a mechanism called *biased competition*. If we are looking for tomatoes, then it is as if an instruction has been issued.

"All you red-sensitive cells in V1, you all have permission to shout louder. All you blue- and green-sensitive cells, try to be quiet." Similar instructions can be issued for particular orientations, or particular sizes—these are all features processed by V1. The responses from the cells that are thereby sensitized are passed both up the *what* pathway, biasing the things that are seen, and up the *where* pathway, to regions that send signals to make eye movements occur. Other areas buried deeper in the brain, such as the hippocampus, are also involved in setting up actions.

In a search for tomatoes all red patches in the visual field of the searcher become candidates for eye movements, and the one that causes red-sensitive neurons to shout the loudest will be visited first. The same biased shouting mechanism also applies to any of the feature types processed by the primary visual cortex, including orientation, size, and motion. The important point here is that knowing what the primary visual cortex does tells us what is easy to find in a visual search.

This is by no means the whole story of eye-movement planning. Prior knowledge about where things are will also determine where we will look. However, in the absence of prior knowledge, understanding what will stand out on a poster, or computer screen, is largely about the kinds of features that are processed early on, before the where and what pathways diverge.

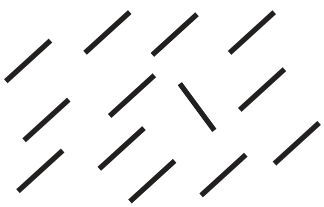
WHAT STANDS OUT = WHAT WE CAN BIAS FOR

Some things seem to *pop out* from the page at the viewer. It seems you could not miss them if you tried. For example, Anne Triesman's name almost certainly popped out at you the moment you turned to this page. Ann Triesman is the psychologist who was the first to systematically study the properties of simple patterns that made them easy to find. Triesman carried out dozens of experiments on our ability to see simple colored shapes as distinct from other shapes surrounding them.◆

◆A large number of popout experiements are summarized in A. Triesman, and S. Gormican, 1988. Feature analysis in early vision: Evidence from search asymmetries, *Psychological Review*. 95(1): 15–98.

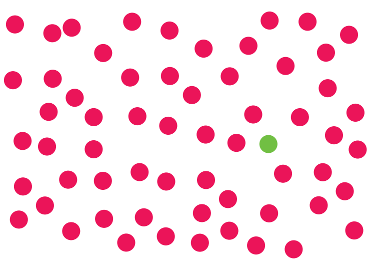
Triesman's experiments consisted of visual search tasks. Subjects were first told what the target shape was going to be and given an instant to look at it; for example, they were told to look for a tilted line. 

Next they were briefly exposed to that same shape embedded in a set of other shapes called distracters.

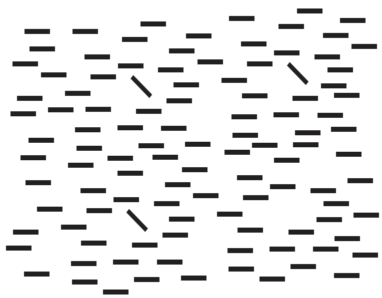


They had to respond by pressing a "yes" button if they saw the shape, and a "no" button if they did not see it. In some trials, the shape was present, and in others it was not.

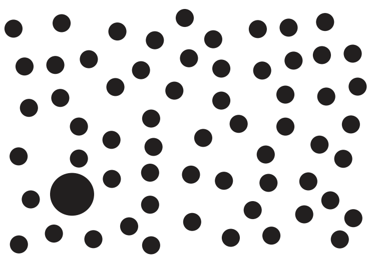
The critical finding was that for certain combinations of targets and distracters the time to respond *did not* depend on the number of distracters. The shape was just as distinct and the response was just as fast if there were a hundred distracters as when there was only one. This suggests a parallel automatic process. Somehow all those hundred things were being eliminated from the search as quickly as one. Triesman claimed that the effects being measured by this method were *pre-attentive*. That is, they occurred because of automatic mechanisms operating prior to the action of attention and taking advantage of the parallel computing of features that occurs in V1 and V2.



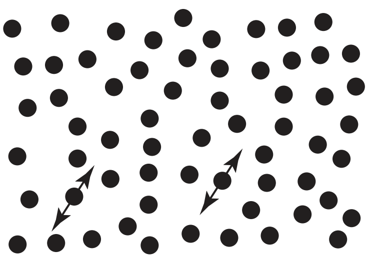
The green dot pops out



The oblique lines pop out



The large circle pops out



If two dots were to oscillate as shown they would pop out

Pop-out effects depend on the relationship of a visual search target to the other objects that surround it. If that target is distinct in some feature channel of the primary visual cortex we can program an eye movement so that it becomes the center of fixation.

Although these studies have contributed enormously to our understanding of early-stage perceptual processes, pre-attentive has turned out to be an unfortunate choice of term. Intense concentrated attention is required for the kinds of experiments Triesman carried out, and her subjects were all required to *focus their attention on the presence or absence of* *a particular target.* They were paid to attend as hard as they could. More recent experiments where subjects were not told of the target ahead of time show that all except the most blatant targets are missed.◆ To be sure, some things shout so loudly that they pop out whether we want them to or not. A bright flashing light is an example. But most of the visual search targets that Triesman used would not have been seen if subjects had not been told what to look for, and this is why pre-attentive is a misnomer.

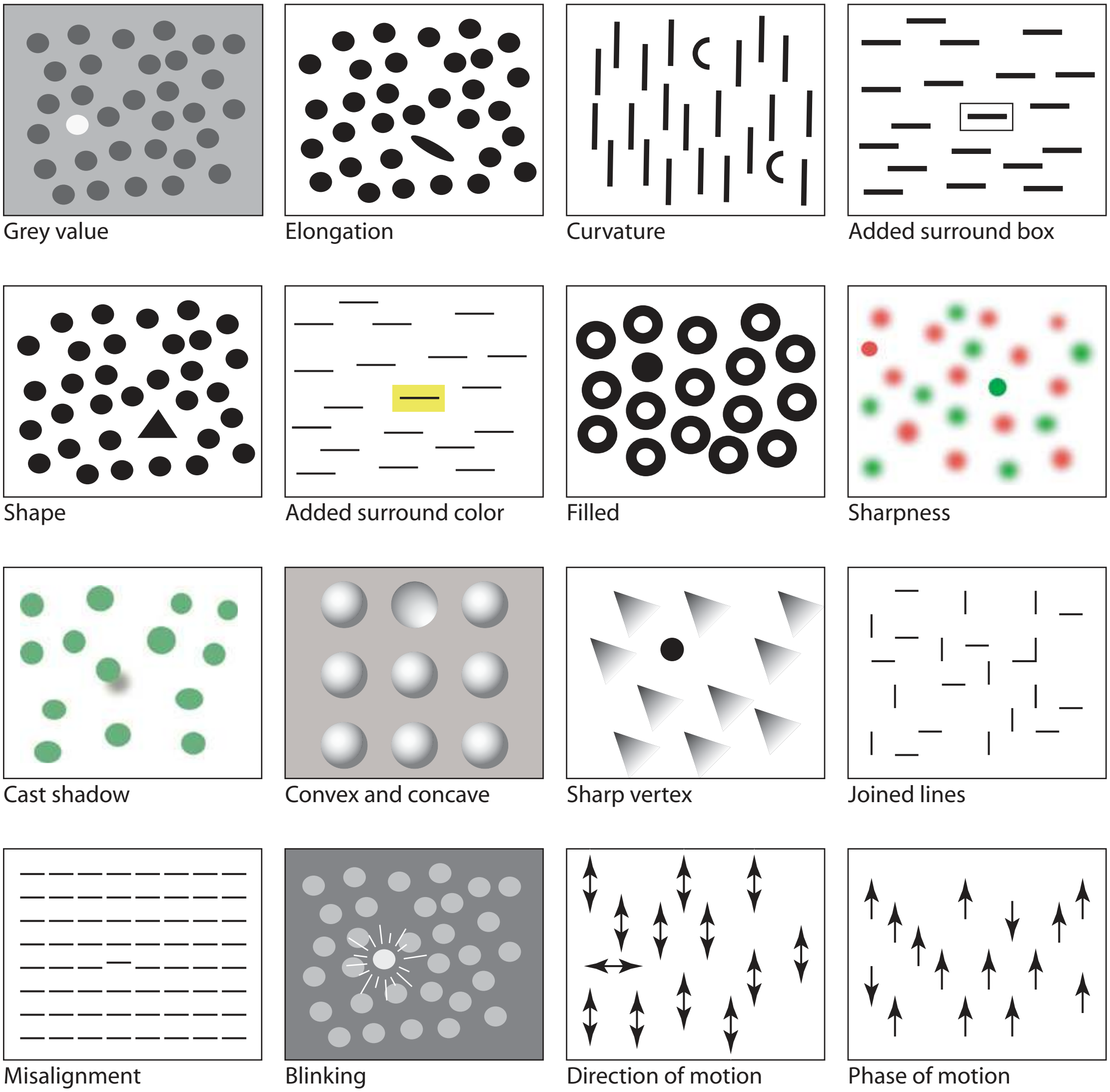
◆The phenomenon of not seeing things that should be obvious is called inattentional blindness. A. Mack and I. Rock, 1998. *Inattentional Blindness*. MIT Press, Cambridge, MA.

A better term would be *tunable,* to indicate those visual properties that can be used in the planning of the next eye movement. Triesman's experiments tell us about those kinds of shapes that have properties to which our eye-movement programming system is sensitive. These are the properties that guide the visual search process and determine what we can easily see.

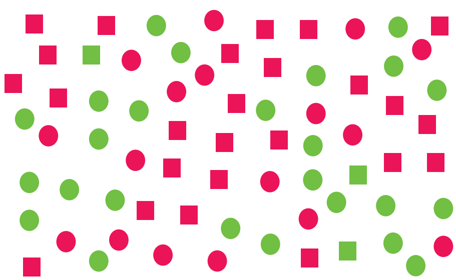
The strongest pop-out effects occur when a single target object differs in some feature from all other objects and where all the other objects are identical, or at least very similar to one another. Visual distinctness has as much to do with the visual characteristics of the *environment* of an object as the characteristics of the object itself. It is the degree of feature-level *contrast* between an object and its surroundings that make it distinct. From the purposes of understanding pop-out, contrast should be defined in terms of the basic features that are processed in the primary visual cortex. The simple features that lead to pop out are color, orientation, size, motion, and stereoscopic depth. There are some exceptions, such as convexity and concavity of contours, that are somewhat mysterious, because primary visual cortex neurons have not yet been found that respond to these properties. But generally there is a striking correspondence between pop-out effects and the early processing mechanisms.

Something that pops out can be seen in a single eye fixation and experiments show that processing to separate a pop-out object from its surroundings actually takes less than a tenth of a second. Things that do not pop out require several eye movements to find, with eye movements taking place at a rate of roughly three per second. Between one and a few seconds may be needed for a search. These may seem like small differences, but they represent the difference between visually efficient at-a-glance processing and cognitively effortful search.

So far we have only looked at patterns that show the pop-out effect. But what patterns *do not* show pop-out? It is equally instructive to examine these. At the bottom of the following page, there is a box containing a number of red and green squares and circles. When you turn the page, look for the green squares.

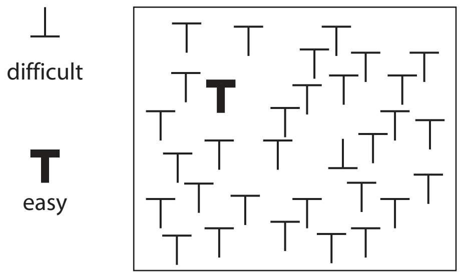


What makes a feature distinct is that it differs from the surrounding feature in terms of the signal it provides to the low-level feature-processing mechanisms of the primary visual cortex.

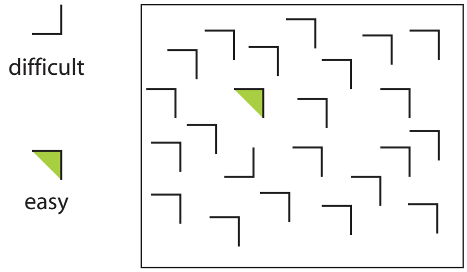


There are three green squares in this pattern. The green squares do not show a pop-out effect, even though you know what to look for. The problem is that your primary visual cortex can either be tuned for the square shapes, or the green things, but not both.

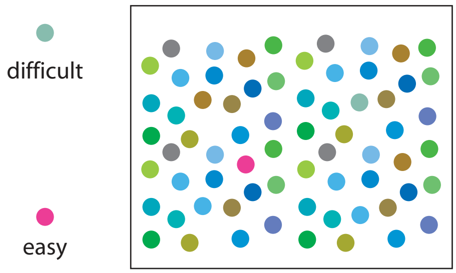
Trying to find a target based on two features is called a visual conjunctive search, and most visual conjunctions are hard to see. Neurons sensitive to more complex conjunction patterns are only found farther up the *what* processing pathway, and these cannot be used to plan eye movements. In each of the following examples, there is something that is easy to find and something that is not easy to find. The easy-to-find things can be differentiated by V1 neurons. The hard-to-find things can only be differentiated by neurons farther up the what pathway.



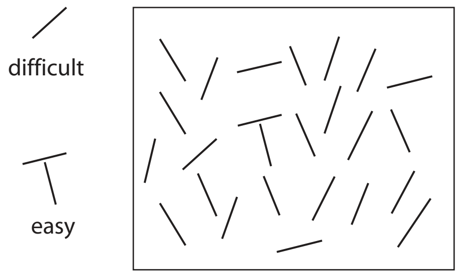
The inverted T has the same feature set as the right-side-up T and is difficult to see. But the bold T does support pop-out and is easy to find.



Similarly the backwards L has the same feature set as the other items, making it difficult to find. But the green triangle addition does pop out.



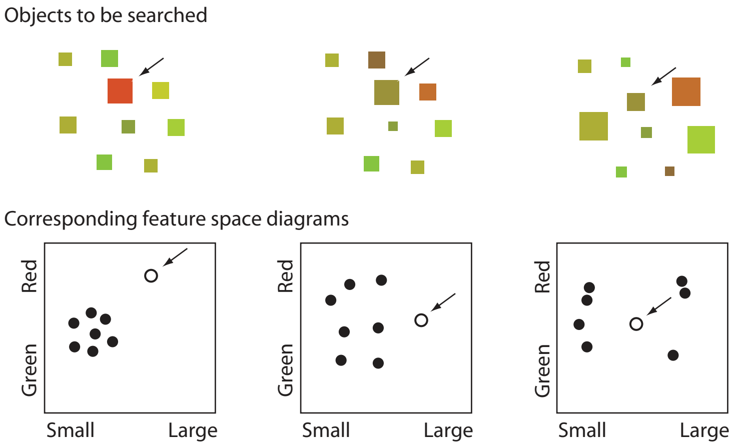
A color that is close to many other similarly colored dots cannot be tuned for and is difficult to find.



Similarly, if a line is surrounded by other lines of various similar orientations it will not stand out.

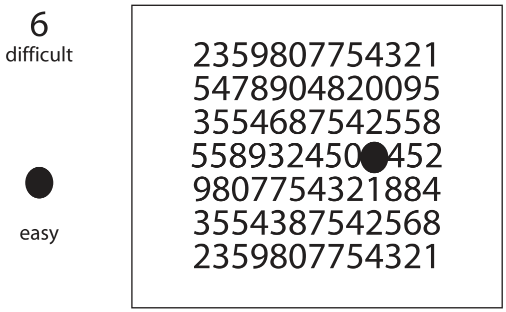
For the pop-out effect to occur, it is not enough that low-level feature differences simply exist, they must also be sufficiently large. For example, as a rule of thumb a thirty-degree orientation difference is needed for a feature to stand out. The extent of variation in the background is also important. If the background is extremely homogenous—for example, a page of twelve-point text—then a small difference is needed to make a particular feature distinct. The more the background varies *in a particular feature channel*—such as color, texture, or orientation—then the larger the difference required to make a feature distinct.

We can think of the problem of seeing an object clearly in terms of *feature channels*. Channels are defined by the different ways the visual image is processed in the primary visual cortex. Feature channels provide a useful way of thinking about what makes something distinct. The diagrams below each represent two feature channels. You can think of one axis representing the color channel and the other axis representing the size channel. The circle represents a target symbol having a particular color and size. This is what is being searched for. The filled circles represent all of the other features on the display.



TOP ROW. Three sets of objects to be searched. In each case an arrow shows the search target. LEFT BOTTOM ROW. The corresponding feature space diagrams. If a target symbol differs on two feature channels it will be more distinct than if it differs only on one. On the left panel the target differs in both color and size from non-targets. In the middle panel the target differs only in size from non-targets. A target will be least distinct if it is completely surrounded in feature space as is shown in the right panel.

One might think that finding things quickly is simply a matter of practice and we could learn to find complex patterns rapidly if we practiced enough. The fact is that learning does not help much. Visual learning is



The number 6 cannot be picked out from all the other numbers despite a lifetime's experience looking at numbers; searching for it will take about two seconds. By contrast the dot can be seen in a single glance, about ten times faster. The features that pop out are hardwired in the brain, not learned.

Individual faces also do not pop out from the crowd even though it may be our brother or sister we are looking for. However, the pinkish round shapes of faces in general may be tunable, and so our visual systems can program a series of eye movements to faces. This does not help much in this picture because there are so many faces. The yellow jacket in the image on the right can be found with a single fixation because there is only one yellow object and no other similar colors.

a valuable skill and with practice experts can interpret patterns that non-experts fail to see. But this expertise applies more to identifying patterns once they have been fixated with the eyes, and not to finding those patterns out of the corner of the eye.

LESSONS FOR DESIGN

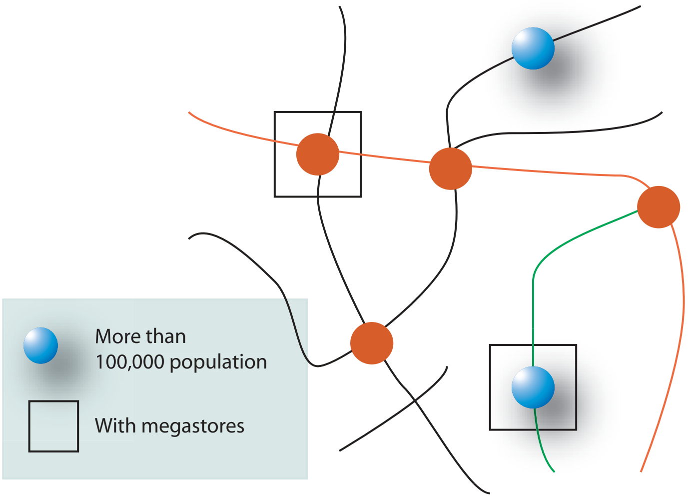
The lessons from these examples are straightforward. If you want to make something easy to find, make it different from its surroundings according to some primary visual channel. Give it a color that is substantially different from all other colors on the page. Give it a size that is substantially different from all other sizes. Make it a curved shape when all other shapes are straight; make it the only thing blinking or moving, and so on.

Many design problems, however, are more complex. What if you wish to make several things easily searchable at the same time? The solution is to use different *channels*. As we have seen, layers in the primary visual cortex are divided up into small areas that *separately* process the elements of form (most importantly, orientation and size), color, and motion. These can be thought of as semi-independent processing channels for visual information.

A design to support a rapid visual query for two different kinds of symbols from among many others will be most effective if each kind of query uses a different channel. Suppose we wish to understand how gross domestic product (GDP) relates to education, population growth rates, and city dwelling for a sample of countries. Only two of these variables can be shown on a conventional scatter plot, but the other can be shown using shape and color. We have plot the (GDP) against population growth rate, and we use shape coding for literacy and color coding for urbanization.

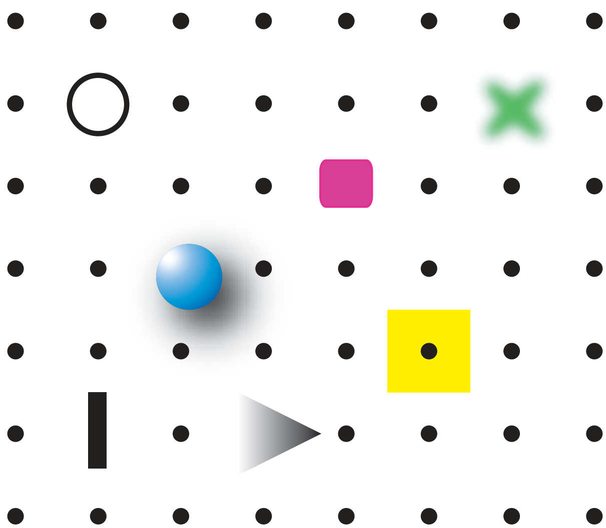
In this scatter plot, two different kinds of points are easy to find. It is easy to visually query those data points representing countries with a high level of literacy. These use color coding. It is also easy to visually query the set of points representing countries with a low urban population. These are distinct on the orientation channel because these symbols are made with **X**s containing strong oblique lines.

It is not necessary to restrict ourselves to a single channel for each kind of symbol. If there are differences between symbols on multiple channels they will be even easier to find. Also, tunability is not an all-or-nothing property of graphic symbols. A symbol can be made to stand out in only a single feature channel. For example, the size channel can be used to make a symbol distinct, but if the symbol can be made to differ from other symbols in both size and color, it will be even more distinct.



Suppose that we wish to display both large towns and towns with megastores on the same map. Multiple tunable differences can be used. The blue symbols are shaded in 3D and have cast shadows. The outline squares are constructed from the only bold straight vertical and horizontal lines on the map. This means that visual queries for either large towns or towns with megastores will be efficiently queried.

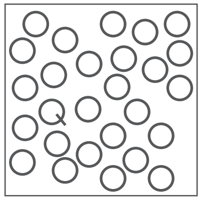
Any complex design will contain a number of background colors, line densities and textures, and the symbols will similarly differ in size, shape, color, and texture. It is a challenge to create a design having more than

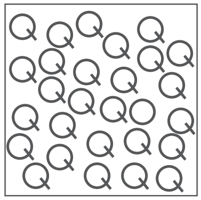


A set of symbols designed so that each would be independently searchable.

Each symbol differs from the others on several channels. For example, there is only one green symbol; it is the only one with oblique lines and it is the only one with no sharp edges.

two or three symbols so that each one will support rapid pop-out searching. Creating a display containing more than eight to ten *independently* searchable symbols is probably impossible simply because there are not enough channels available. When we are aiming for pop-out, we only have about three difference steps available on each channel: three sizes, three orientations, three frequencies of motion, etc. The following set is an attempt to produce seven symbols that are as distinct as possible given that the display is static.





There is an inherent tradeoff between stylistic consistency and overall clarity that every designer must come to terms with. The seven different symbols are stylistically very different from each other. It is easy to construct a stylistically consistent symbol set using either seven different colors or seven different shapes, but there is a cognitive cost in doing so. Visual searches will take longer if only color coding is used, than when shape, texture, and color differentiate the symbols, and in a complex display the difference can be extreme. What takes a fraction of a second with the multi-feature design might easily take several seconds with the consistent color-only design.

Many kinds of visibility enhancements are not symmetric. Increasing the size of a symbol will result in something that is more distinctive than a corresponding decrease in size. Similarly, an increase in contrast will be more distinctive than a decrease in contrast. Adding an extra part to a symbol is more distinctive than taking a part away.

There is a kind of visual competition in the street as signs compete for our attention.

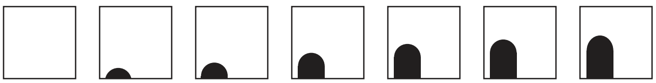
Blinking signs are the most effective in capturing our attention. When regulation allows it and budgets are unlimited, Times Square is the result.

MOTION

Making objects move is a method of visibility enhancement that is in a class by itself. If you are resting on the African Savannah, it pays to be sensitive to motion at the edge of your visual field. That shaking of a bush seen out of the corner of your eye might be the only warning you get that something is stalking you with lunch in mind. The need to detect predators has been a constant factor through hundreds of millions of years of evolution, which presumably explains why we and other animals are extremely sensitive to motion in the periphery of our visual field. Our sensitivity to *static* detail falls off very rapidly away from the central fovea. Our sensitivity to motion falls off much less, so we can still see that something is moving out of the corner of our eye, even though the shape is invisible.

Motion is extremely powerful in generating an *orienting response*. It is hard to resist looking at an icon jiggling on a web page, which is exactly why moving icons can be irritating. A study by A.P. Hillstrom and S. Yantis suggests that the things that most powerfully elicit the orienting response are not simply things that move, but things that *emerge* into the visual field.◆ We rapidly become habituated to simple motion; otherwise, every blade of grass moving would startle us.

◆A.P. Hillstrom and S. Yantis, 1994. Visual attention and motion capture. *Perception and Psychophysics*. 55(4): 109–154.



In the design of computer interfaces, one good use of motion is as a kind of human interrupt. Sometimes people wish to be alerted to incoming email or instant messaging requests. Perhaps in the future virtual agents will scour the Internet for information and occasionally report back their findings. Moving icons can signal their arrival. If the motion is rapid, the effect may be irritating and hard to ignore, and this would be useful for urgent messages. If the motion is slower and smoother, the effect can be a gentler reminder that there is something needing attention. Of course, if we want to use the emergence of an object to provoke an orienting response on the part of the computer user it is not enough to have something emerge only once. We do not see things that change when we are in the midst of making an eye movement or when we are concentrating. Signaling icons should emerge then disappear every few seconds or minutes to reduce habituation.●

●I introduced the idea of using motion as a human interrupt in a paper I wrote with collaborators in 1992. C. Ware, J. Bonner, W. Knight and R. Cater. Moving icons as a human interrupt. *International Journal of Human–Computer Interaction*. 4(4): 173–178.

The web designer now has the ability to create web pages that crawl, jiggle, and flash. Unsurprisingly, because it is difficult for people to suppress the orienting response to motion, this has provoked a strong aversion among users to the web sites where these effects flourish. The gratuitous use of motion is one of the worst forms of visual pollution, but carefully applied motion can be a useful technique. This is not to say that motion per se is bad. We can be outdoors where trees sway, clouds move, and people pass to and fro without feeling irritated. It is especially high-frequency rapid motion or blinking that induces the unavoidable orienting response. Another caveat with respect to motion is that many people like the energy and stimulation that motion imparts—one person's amusement arcade is another person's hell. We should make the distinction between situations where the goal is entertainment and situations where the goal is providing information in the most effective possible way.

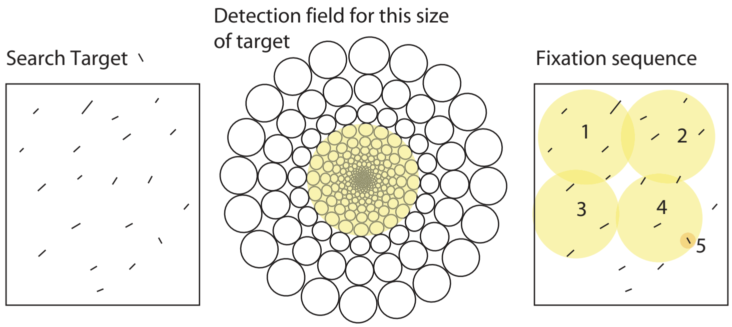
VISUAL SEARCH STRATEGIES AND SKILLS

So far we have focused on the specific feature properties that are used in the very short-term planning of the next eye movement, but this is by no means the whole story. Finding a small object is a skilled strategic process. At every instant in time part of the brain is planning the next eye movement based on the information available from the current fixation of the eye as well as some small amount that is retained from the previous few fixations. In addition, our previous experiences with similar visual environments also influence the visual search pattern and this is where skill comes in. Our prior experience tells us where to look for something, and this information is stored as a pattern of eye movement tendencies. When we read a newspaper, for example, we are likely to scan the main headlines and images first. But the patterns of eye movements that occur in response to a particular scene are never inevitable; it is a fluid just-in-time process, and the potential search targets are continually being reassessed.

A graphic design that has visual structure at several scales can aid the search process. Large-scale structure is needed as a means for finding important mid-scale and small-scale information.

THE DETECTION FIELD

Remember that the size of processing units (brain pixels) increases with distance from the center of gaze (see Chapter 1). Small features cannot be detected by large brain pixels at the edge of visual space. If the targets are small and near the current center of fixation, the brain can apply the selective feature-based biasing mechanism we have been discussing. But if small targets are off to the side, some other process must kick in. Several eye movements will be required to scan the visual field until one lands close enough to the target.



Brain pixels get larger the farther away they are from the point of fixation, thus several eye movements are needed before the target can be detected. On the fourth fixation, the target is detected at the edge of the tunable region. A final eye movement brings it to the center of the fovea.

The area around the center of the fovea where the presence of a particular target may be detected can be thought of as a *detection field*. If a search target is within the detection field, this does not necessarily mean that it can be positively identified; it only means that this can become a target for the next fixation, and after that it may be identified. In the absence of a candidate target in the detection field, the purpose of a *scanning strategy* is to get the eye *in the vicinity* of the target so that the feature-based pop-out mechanism can function as a final step.

A typical scanning strategy is the one we use when reading; we start in the upper-left-hand corner and move our eyes across from left to right starting at the top and progressing down. In most real-world search tasks, the visual field is not uniformly filled with potential candidates. Sometimes the likely candidates are grouped in clusters, and our search strategy will take advantage of this. Big patterns have a much bigger detection field, and this allows for a hierarchical search strategy. If small target patterns are embedded in specific types of large patterns, then this information can be used. First we make an eye movement to the likely neighborhood of a target, based on the limited information in our peripheral vision; next, the local pattern information provides a few candidates for individual detailed fixations.

Our everyday environment has structure at all scales. Rooms contain large furniture objects such as desks, chairs, bookshelves, and cupboards, as well as small hand-sized objects such as books, telephones, and coffee cups. These are arranged in predictable relationships; the books are on the shelves and the desktops, and the coffee cup is on the desk. This allows for a much more efficient visual search. For the thief in the night, the big patterns may be dressers and side tables. The small patterns may be pieces of jewelry. Sometimes it can be enough that the small targets are arranged in clumps, enabling a strategy based on looking first for the clumps, then searching within each clump.

In this example, the targets are clustered and a two-stage search strategy is used. The first stage involves tuning for and making an eye movement to a cluster of targets (seen as fuzzy blobs in the periphery). After the fixation is made, the blob is resolved into individual targets. The second stage involves tuning for and making an eye movement to a particular candidate target within a blob. The final tuning is based on orientation.

The process we have described so far is missing one essential property. How does the search control system avoid revisiting places that have already been examined? Unless there is a memory for where eye fixations have recently been placed, the point just visited will be the best candidate for the next fixation. Without a blocking mechanism the eye would be trapped, flicking back and forth between the two most likely target areas. The brain must somehow mark the locations of recent fixations and inhibit the tendency to revisit them. It is thought that a structure on the *where* pathway called the *lateral interparietal* area performs this function.◆ Experimental evidence suggests that between four and six locations recently visited with eye movements are retained. In some cases, the identity of the object at a particular location may also be retained in visual working memory, but in other cases the brain simply flags the fact that a particular location has been visited.

◆See James W. Bisley, and Michael E. Goldberg, 2003. Neuronal activity in the lateral interparietal area and spatial attention. *Science*. 299: 81–86.

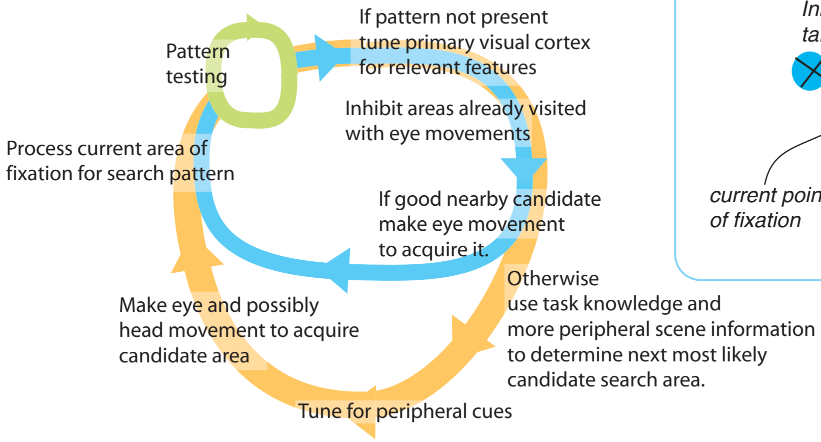
THE VISUAL SEARCH PROCESS

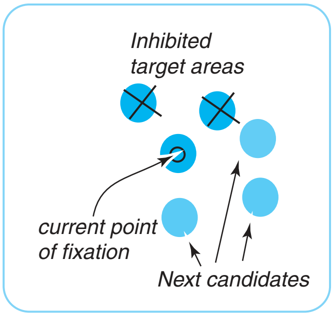
We can now elaborate on the inner loops of visual problem solving that were introduced in Chapter 1, concentrating on the eye movement control process. We will start with the outermost loop and work in.

Move and scan loop. Assuming that we know what we are looking for, when we enter an environment, our initial search strategy will involve orienting the head and perhaps walking to get the best viewpoint. From this vantage, we will initiate a sequence of fixations. If the target is not found we move to a new vantage point to continue the search.

Eye movement control loop. Planning and executing eye movements occurs between one and three times per second. This involves both the biasing mechanism, so that new candidate targets can be determined based on their elementary properties of form (orientation, size, color, motion), and a simple map of what regions have been recently visited by means of eye movements.

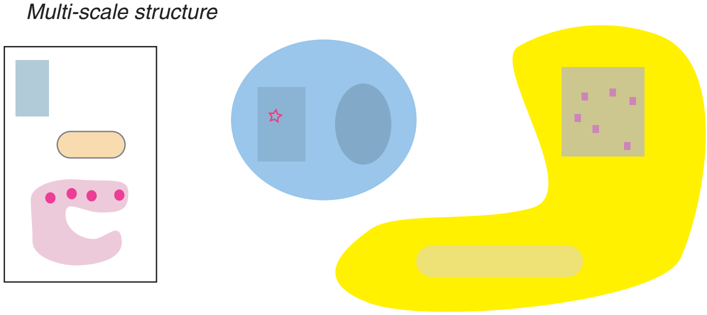
Pattern-testing loop. When the eye alights on a promising target area, the inner loop function is executed. This involves testing the pattern to see if it is the search target or not. The brain takes about one-twentieth of a second to make each test; typically between one and four tests are made on each fixation.





USING MULTISCALE STRUCTURE TO DESIGN FOR SEARCH

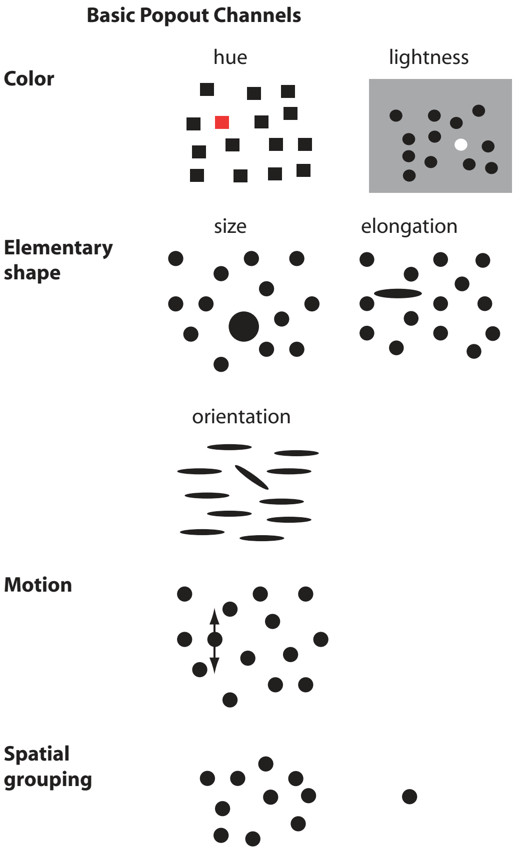
To support efficient visual search, a design should be given large-scale as well as small scale structure. All too often, artificial information displays lack visual structure at different scales. Menus and windows tend to all look the same. One rectangular box is much like another. Adding multiscale visual structure will make search much more efficient, as long as the smaller objects of search can predictably be associated with larger visual objects.



Having structure at multiple scales is most important for designs that will be used over and over again. It permits visual search skills to develop in the form of eye movement sequences that occur in response to the general properties of a particular scene. But even for designs that are used only for a few minutes, high-level structure supports location memory and makes it easier to revisit places that have been looked at only seconds ago.

CONCLUSION

Visual search is not an occasional activity only occurring when we have lost something. Because we have so little information in our heads, search is something that is fundamental to almost all seeing. Even though we are mostly unaware that we are doing it, the visual world is being reassessed in terms of where we should look next on every fixation. Two seconds is a long time in a visual thinking. There is a world of difference between something that can be located with a single eye movement and one that takes five or ten. In the former case, visual thinking will be fluid, in the latter, it will be inefficient and frustrating.



One key to making efficient visual search is through the use of pop-out properties. If a visual object is distinct on one or more of the visual channels, then it can be processed to direct an eye movement. The strongest pop-out differentiators are the basic feature channels found in V1. These include the elements of form, size, elongation, and orientation; the elements of color, including hue and lightness (these are discussed further in Chapter 4); and motion, and spatial layout. The figure on the previous page gives a summary.

Large-scale graphic structure can also help with visual search, but only if the searcher *already knows* where in a large structure an important detail exists. Such knowledge is built up as we scan an image and graphic structure is always helpful in enabling us to return to something we have recently seen. Knowledge also comes, however, from our previous experience with similar graphics and this is a two-edged sword. If the design being searched conforms to stereotype then search will be easy because habitual visual scan patterns, guided by spatial structure, will support search. If a design violates a learned pattern then searches using habitual eye movement strategies will result in frustration.