Visual search

Jeremy M. Wolfe

Visual Attention Laboratory, Brigham & Women's Hospital, 64 Sidney Street Suite 170, Cambridge, MA 02139-4170, USA. wolfe@search.bwh.harvard.edu

In Shakespeare's *Henry IV Part 1*, the Scottish rebel, the Earl of Douglas, engages in a visual search task. He is searching for King Henry in a field full of soldiers who are not King Henry (known in the search trade as 'distractors'). The problem is that some of those soldiers are wearing the colors of the king. As his ally, Hotspur, puts it, "The king hath many marching in his coats". Douglas proposes what would be considered a serial, self-terminating search strategy for completing this task: "Now, by my sword, I will kill all his coats; I'll murder all his wardrobe, piece by piece, until I meet the king". Typical searches in the lab and in everyday life are of a less sanguinary nature. Nevertheless, we can detect a variety of principles in the Earl of Douglas' search that are common to searches for a red X on a computer screen, the salt shaker on the dinner table, or a threat in an X-ray of carry-on luggage.

To begin, search is necessary, even for targets that are in plain view. In Figure 1, the King of Diamonds is the target. Even though those targets are visible and even though you have an impression of a field full of identifiable playing cards, you can only confirm the presence of the specific target once you succeed in directing attention to the target item. It is simply not possible to fully process all of the stimuli in the visual field at one time. By the way, if you only found one King of Diamonds, try to find the other one! In the radiology literature, a failure to find all of the targets is known as 'satisfaction of search'.

Douglas has no intention of swinging his sword randomly because he can use information about the appearance of the royal coat to guide his attention. The deployment of attention in any search is guided by one or more sources of information. The most extensively studied of these sources of guidance are the so-called preattentive attributes, a limited set of attributes such as color, motion, size, and so forth that can be processed across the visual field in parallel — that is, across the whole field in a single step. You do not need to select a specific object or location in order to determine that there is redness, motion, and so on at that location. Signals derived from that preattentive processing can be used to guide attention in two ways. First, attention is summoned in a bottom-up, stimulus-driven manner to the most salient item(s) in a display. In Figure 2, it is likely that, in the absence of other instruction, your attention was attracted to the large, irregular, yellow item in this bottom-up manner. Second, attention can be guided in a top-down, user-driven manner. If you search Figure 2 for the small, oblique, blue oval, you will be able to use all four of those features to guide your attention. There are probably one to two dozen guiding attributes. Some of these, color for example, are uncontroversial; others, faces for example, have generated substantial literatures without firmly establishing themselves as preattentive attributes.

The signals that are used to guide attention are clearly related to what we mean when we colloquially refer to color or orientation or some other basic attribute of vision, but guidance by an attribute can differ from perception of that attribute. For example, suppose we find the desaturated colors that lie perceptually half way between white and saturated hues (for example, a light green, halfway between saturated green and white). It turns out that the reddish/pinkish hues are found much more quickly in a search than are light greens, blues, or purples, which are equally distinctive perceptually. Apparently, color information is packaged one way for perceptual experience and another way for attentional guidance.

Top-down and bottom-up guidance can work against each other. Over the years, considerable research has been devoted to studying the degree to which a task-irrelevant item can capture attention. Thus, if you had been asked to find red ovals and are then presented with Figure 2, you might have found your attention captured by the irregular yellow item even though it is neither red nor oval. If your attention was captured, how does it get free? It is likely that attended and rejected items are inhibited in a manner that allows attention to escape and produces a bias against selecting that item again, at least for a while (a phenomenon known as 'inhibition of return'). The observer's history can influence what captures attention. For example, if the target happened to be red on the previous trial, attention would be more likely to be drawn to a red item, even an irrelevant red item, in the current search display. The previous trial has primed the current trial, in effect making the red items more salient. This is known as 'priming of pop-out'.

The slope of the function relating reaction time to set size is a useful metric for measuring the extent of guidance in a search. 'Reaction time' is the time required to say "yes", there is a target present or "no", there is not. It could also be the time to localize the target or to make some other response like categorizing the target's shape. 'Set size' simply refers to the number of items in the display (sixteen in Figure 1). The slope measures the rate at which items can be processed in a search. It is not a measure of the time required for each item to be processed unless one can be sure that only one item is being processed at a time. Thus, if each item needs to be fixated, requiring an eye movement from one item to the next, the rate will be limited by the rate of those eye movements, which is typically three or four per second. A search among large, well-separated objects would not be limited by eye movements. If all items in that display were preattentively equivalent, the resulting unguided search would tend to have a slope in the range of 20–40 milliseconds per item for target-present trials and a bit more than twice that for trials when the target is absent. Translating those slopes into an estimate of items per second requires assumptions about the nature of search. For example, how much memory is there for the history of the search? Are items sampled with or without replacement?

If guidance is perfect, then attention can be directed to the target on its first deployment, regardless of the number of distractors. Under these conditions of 'pop-out', the slope will be near zero. Intermediate cases where there is some guidance (as in Figures 1 and 2) will tend to produce intermediate slopes. The Earl of Douglas is performing one of these intermediate searches. Many objects in his visual field can be readily eliminated from the set of candidate targets, but a subset remain as possible kings who apparently must be murdered piece by piece until he finds the king.

The astute reader will have noticed some differences between our Shakespearean search example and the searches that have provided the scientific data. The bulk of the work described thus far has involved asking observers to search through arrays of simple items, sometimes a very few items, presented in isolation on an otherwise blank background on the computer screen. The Earl of Douglas, in contrast, was searching the battlefield at Shrewsbury. Search in real scenes introduces factors not present in searches of isolated items. Thus, if asked to find a black cow on the now-peaceful field at Shrewsbury (Figure 3), your attention would be guided by aforementioned features such as color, but also by knowledge of properties of the physical world. Cows, under most circumstances, will be found on the ground plane and not in the sky or on vertical walls. Research on guidance of this sort is in a comparatively early stage.

It is clear that some aspects of scene structure are extracted very rapidly, probably without the need for selective attention. Similarly, some semantic information, such as the knowledge that this is an outdoor scene, can be computed even if attention is occupied elsewhere (though it is notoriously difficult to prove that no attention leaked into a task like scene classification). Other matters are less clear. How efficient is search in scenes? The standard measure, the slope of a reaction time by set size function, would require a measure of the set size of the scene. What is the set size in Shrewsbury field? It is hard to know. Is a cow an object? What about its head? You certainly could search for a white head. What is the effect of the cluttered nature of the real world? You must have some internal representation of the object of search. What is the internal representation of a search target like 'King Henry' or 'white cow head', given that they can be recognized in a substantial array of positions, orientations, and distances?

Another complication of the real world is that stimuli can be ambiguous. It is not always obvious if the current object of attention is, or is not, the target. In the realm of academic research, this has been studied with the tools of signal detection theory, using simple stimuli presented too briefly to permit perfect identification. In the real world, this is an important issue in tasks like airport security and cancer screening (Figure 4), even when the stimuli are visible for an extended period of time. Even well-trained experts make a substantial number of errors in these difficult tasks. Unless and until improvements in technology can eliminate errors, it will be important to understand the sources of misses and false alarms and the forces that drive observers toward one or the other type of error. Cancer and airport screening are searches for very rare targets. In the lab, this biases searchers to say "no" and to miss a larger number of targets. Miss errors that could lead to a cancer death or terrorist attack are far more costly than false alarms in these settings, though the false alarms also come with costs, financial and emotional. In the lab, the high cost of misses would push observers toward a more liberal decision criterion and more false alarm errors. Now factor in the time pressure imposed by the line in Terminal C or by the stack of cases waiting to be read and you have a very complex combination of search and signal detection that is being played for very high stakes.

If the target is potentially ambiguous, so too are the distractors. When stimuli are flashed briefly, this makes the decision problem into a task of deciding if any of N items produces enough of a signal to be deemed to be a target. For more extended stimuli, especially for

those that may need to be attended in series, this makes search into a succession of decisions.

Mammography and airport security, to say nothing of searches for English monarchs, raise another disconnect between what we know from the lab and what we want to understand in the world. Many of these real world searches extend over many seconds or minutes (Figure 5). Experimental work is mostly concerned with searches that are completed in a few seconds, at most, and with search stimuli that may be visible for a small fraction of a second. Researchers are not ignoring reality. It is simply impractical to run hundreds of trials that take minutes each. Nevertheless, it is interesting to think about the issues raised when the searcher has been immersed in a scene for some time. Intuitively, it is clear that knowing something about the current visual world aids search. You will find the paper towels more quickly in your own kitchen than in another kitchen. Moreover, because of your knowledge of kitchens, you will find those towels more quickly in that other kitchen than in the basement. Intuition notwithstanding, in the lab, the efficiency of simple but inefficient search does not improve even after hundreds of trials. If you are looking through a set of six letters and, on each trial, you are asked about a different letter that might or might not be present, the third search looks a lot like the thirtieth and the three hundredth. In those simple cases, doing even an inefficient visual search from scratch is probably more efficient than pulling the location of the specific target letter out of memory. In an extended, complex search, however, remembered information about the scene will have the time to become useful. Given that working memory has a very limited capacity, it will be interesting to discover how that capacity is used to support extended search.

Douglas is looking for a single target. This has been true in the bulk of the search literature as well but, of course, that need not be the case. Returning to the photograph of Shrewsbury field, there might be blackberries in that hedge. Finding them involves a visual search, guided by color, shape, texture, and no doubt, by some knowledge of the ways that blackberries insinuate themselves into hedges. This is a search for multiple targets. Given a long hedge with an effectively unlimited set of target berries and non-berry distractors, how long should you search in one spot before moving on? There are proposed answers in the foraging theory and data, largely based on research with animal populations but very little human search data. Our behavior in such situations may have evolved over millennia of hunting and gathering. It remains relevant today when we ask a radiologist, for example, to find all the abnormalities in a series of images from one patient or when we ask an image analyst to mark every interesting finding in a satellite image of some part of a foreign country. In principle, one could search forever, but there is another case, another image, another berry bush waiting. There are theoretical answers to the question of when it is time to move on, but, again, little human data.

In summary, we know a great deal about search but mostly at the short time scale on the continuum illustrated here (Figure 5). There is a rich world of search behavior still to investigate. Oh, and the Earl of Douglas ends up as a prisoner of the future Henry the Fifth. Perhaps an early case of 'attentional capture'.

Further reading

1. Henderson JM. Regarding scenes. Curr. Dir. Psychol. Sci. 2007; 16:219–222.

2. Kristjánsson A, Campana G. Where perception meets memory: a review of repetition priming in visual search tasks. Attent. Percept. Psychophys. 2010; 72:5–18.

3. Muller, HJ., Krummenacher, J., editors. Visual Search and Attention (Special Issue of Visual Cognition, Vol. 14, issues 4-8). Colchester, Essex, UK: Psychology Press, Taylor and Francis Group; 2006.

4. Palmer J, Verghese P, Pavel M. The psychophysics of visual search. Vision Res. 2000; 40:1227–1268. [PubMed: 10788638]

5. Wang Z, Klein RM. Searching for inhibition of return in visual search: a review. Vision Res. 2010; 50:220–228. [PubMed: 19932128]

6. Wolfe JM, Horowitz TS. What attributes guide the deployment of visual attention and how do they do it? Nat. Rev. Neurosci. 2004; 5:495–501. [PubMed: 15152199]
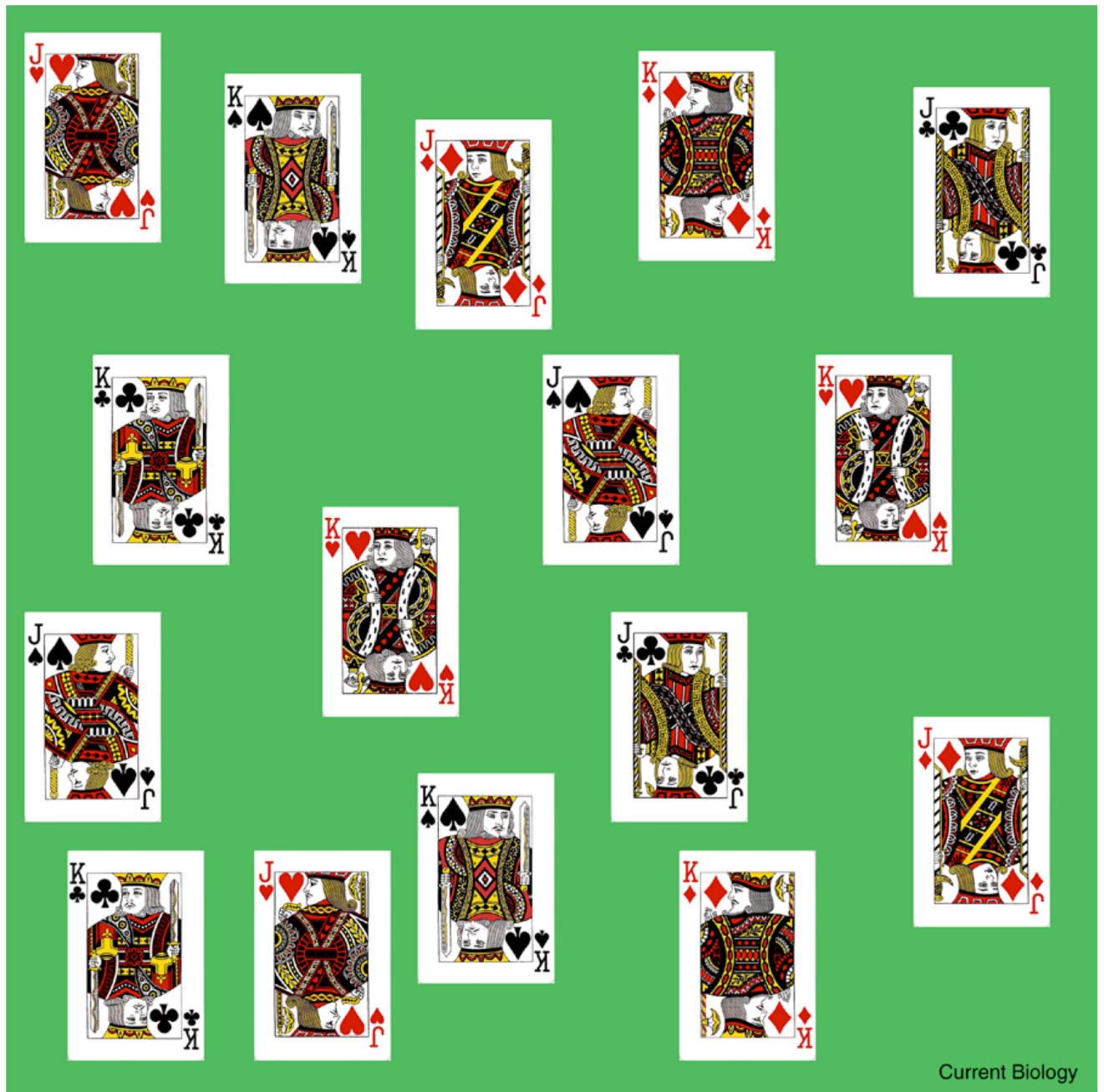
Figure 1.

Search for the King of Diamonds.

Even though you can see all the cards, you need to search for the target. You probably do not spend much time on the black cards.
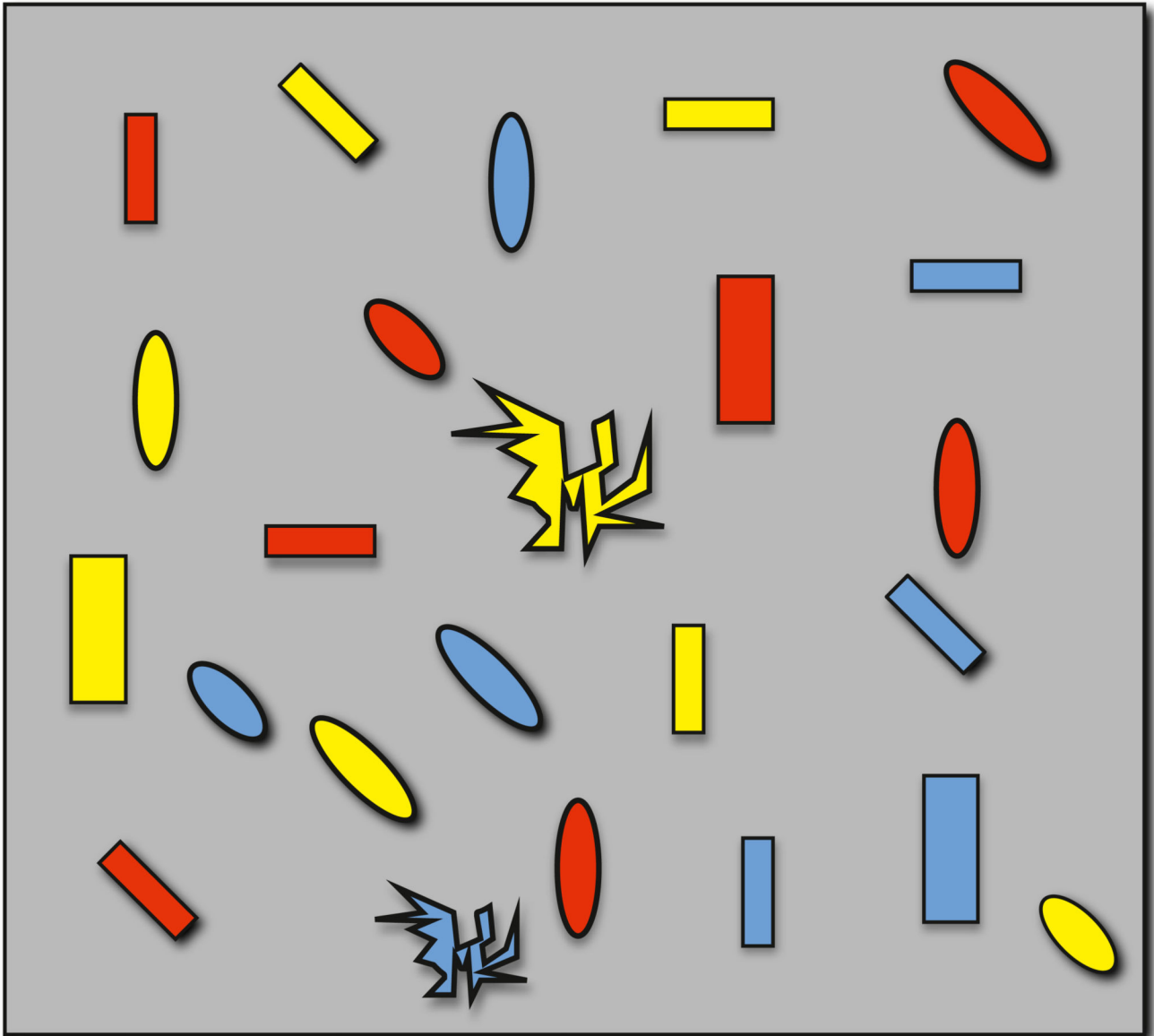
Figure 2.
Attention and visual search.

Your attention is probably drawn to the big, irregular yellow item in a bottom-up, stimulus driven manner. With no change in the visual stimulation, you can easily reconfigure your attention to favor 'blue ovals' in a 'top-down' user-driven manner.

Figure 3.
Most search takes place in the world of continuous scenes, in this case, the now peaceful battlefield at Shrewsbury. (Photo: Marian Byrne, http://www.flickr.com/photos/wonkyknee/3634944485/)
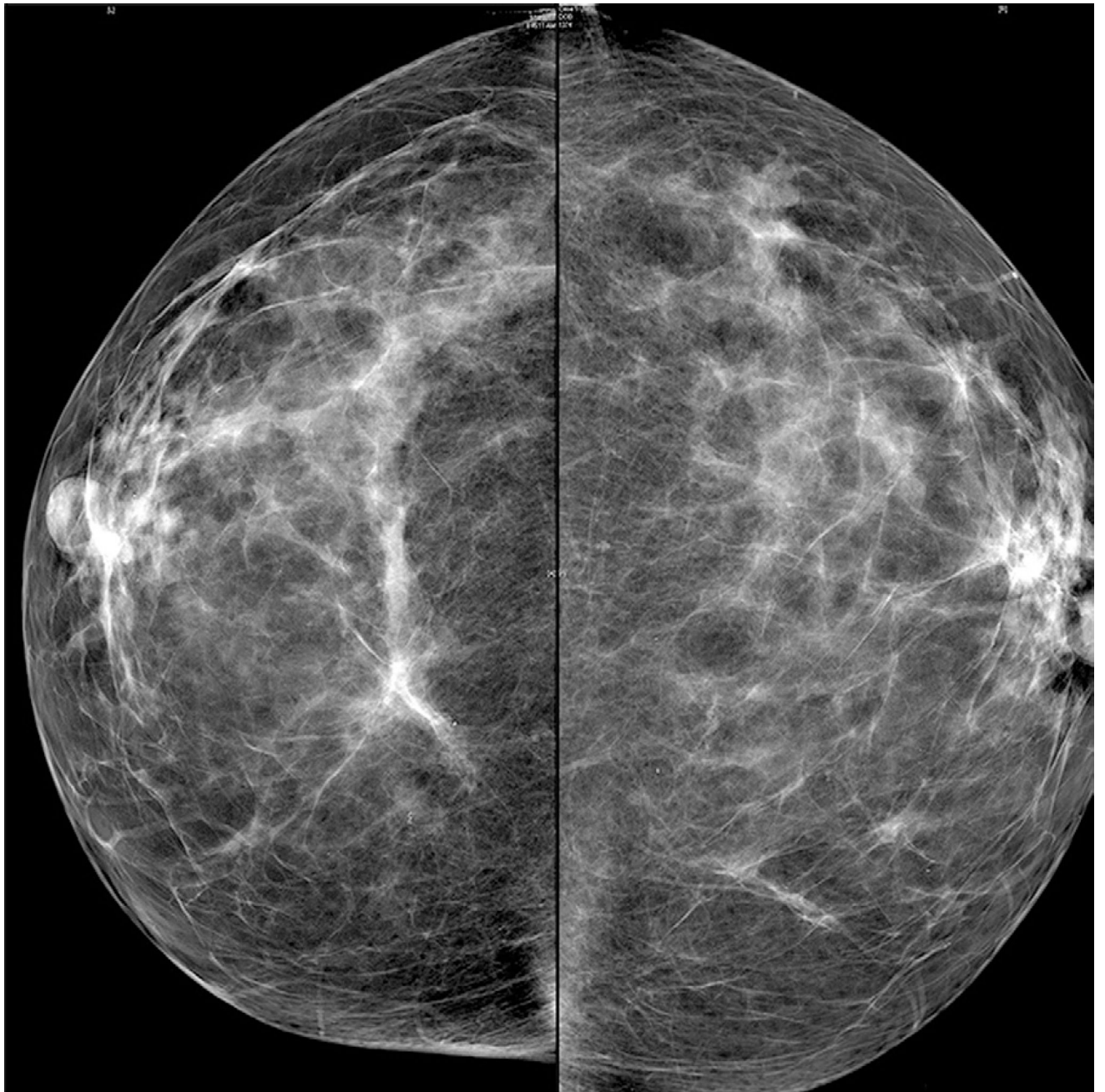
Figure 4.
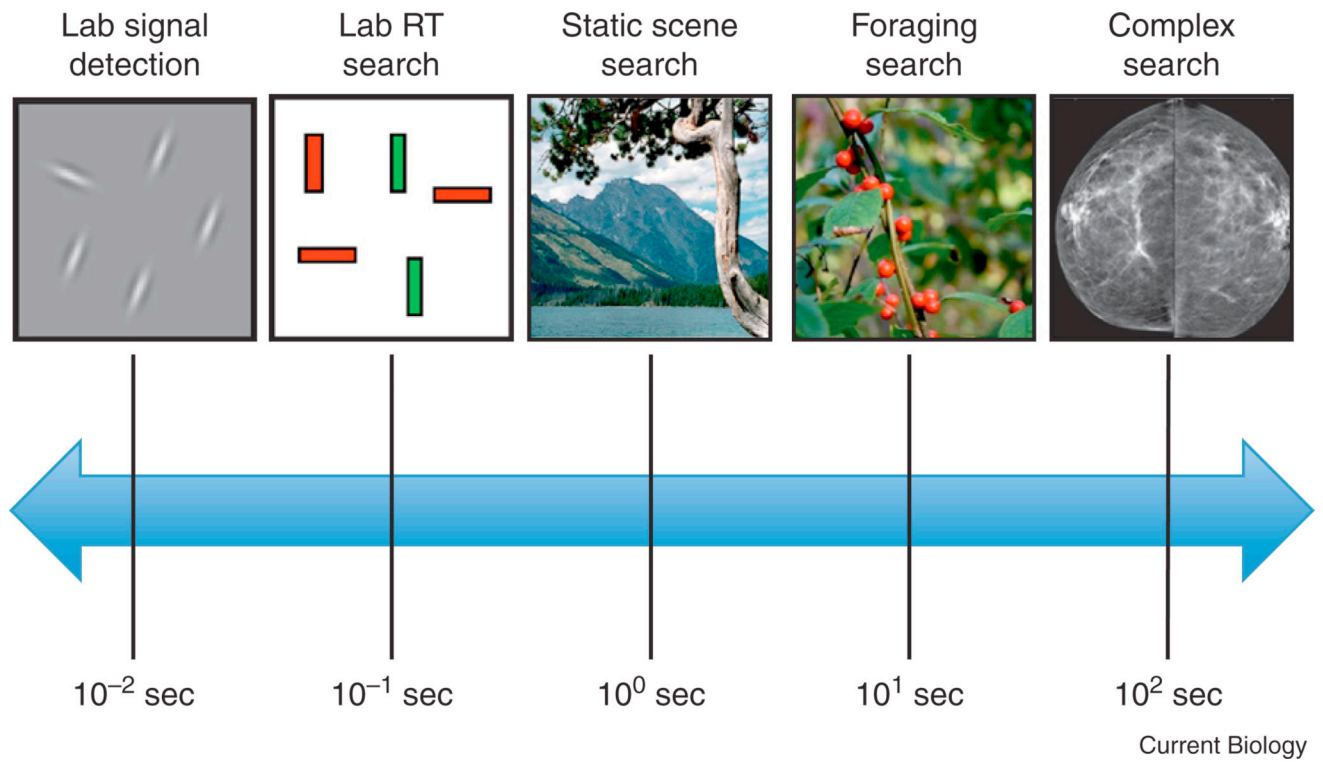Finding signs of cancer in a mammogram can be a difficult visual search task, even for experts.

Figure 5.
Timescale of visual search.

The timescale of typical search tasks runs from fractions of a second to many minutes or more. Most laboratory research has been concentrated at the shorter end of this spectrum.