Visual search in scenes involves selective and non-selective pathways

Jeremy M Wolfe, **Melissa L-H Vo**, **Karla K Evans**, and **Michelle R Greene**
Brigham & Women's Hospital, Harvard Medical School, 64 Sidney St. Suite 170, Cambridge, MA 02139

Abstract

How do we find objects in scenes? For decades, visual search models have been built on experiments in which observers search for targets, presented among distractor items, isolated and randomly arranged on blank backgrounds. Are these models relevant to search in continuous scenes? This paper argues that the mechanisms that govern artificial, laboratory search tasks do play a role in visual search in scenes. However, scene-based information is used to guide search in ways that had no place in earlier models. Search in scenes may be best explained by a dual-path model: A "selective" path in which candidate objects must be individually selected for recognition and a "non-selective" path in which information can be extracted from global / statistical information.

Searching and experiencing a scene

It is an interesting aspect of visual experience that we can look for an object that is, literally, right in front of our eyes, yet not find it for an appreciable period of time. It is clear that we are seeing something at the object's location before we find it. What is that something and how do we go about finding that desired object? These questions have occupied visual search researchers for decades. While visual search papers have conventionally described search as an important real-world task, the bulk of research had observers looking for targets among some number of distractor items, all presented in random configurations on otherwise blank backgrounds. In the last decade, there has been a surge of work using more naturalistic scenes as stimuli and this has raised the issue of the relationship of the search to the structure of the scene. This paper will briefly summarize some of the models and solutions developed with artificial stimuli and then describe what happens when these ideas confront search in real-world scenes. We will argue that the process of object recognition, required for most search tasks, involves the selection of individual candidate objects because all objects cannot be recognized at once. At the same time, the experience of a continuous visual field tell us that some aspects of a scene reach awareness without being limited by the selection bottleneck in object recognition. Work in the past decade has revealed how this non-selective processing is put to use when we search in real scenes.

Classic Guided Search

One approach to search, developed from studies of simple stimuli randomly placed on blank backgrounds can be called "classic Guided Search" [1]. It has roots in Treisman's Feature Integration Theory [2]. As will be briefly reviewed below, it holds that search is necessary because object recognition processes are limited to one or, perhaps, a very few objects at one time. The selection of candidate objects for subsequent recognition is guided by preattentively acquired information about a limited set of attributes like color, orientation, and size.

Object recognition is capacity limited

We need to search because, while we are very good at recognizing objects, we cannot recognize multiple objects at the same time. For example, all of the objects in Figure 1 are simple in construction, but if you are asked to find "T"s that are purple and green, you will find that you need to scrutinize each item until you stumble upon the targets (there are four). It is introspectively obvious that you can see a set of items and could give reasonable estimates for their number, color, and so forth. However, recognition of a specific type of item requires another step of binding the visual features together [3]. That step is capacity-limited and very often, attention demanding [4] (however see [5]) .

In the case of Figure 1, the ability to recognize one object is also going to be limited by the proximity of other, similar items. These "crowding" phenomena have attracted increasing interest in the past few years ([6] [7]). However, though it would be a less compelling demonstration, it would still be necessary to attend to item after item in order to bind their features and recognize them even if there were very few items and even if those items were widely spaced [8].

The selection mechanism is a serial / parallel hybrid

While it is clear that object recognition is capacity-limited, the nature of that limitation has been less clear (for an earlier discussion of this issue, see [9]). The classic debate has been between "serial" models that propose that items are processed one after the other [2] and "parallel" models that hold that multiple objects, perhaps all objects, are processed at the same time but that the efficiency of processing of any one item decreases as the number of items increases [10] [11]. The debate has been complicated by the fact that the classic reaction time data, used in many experiments, is ambiguous in the sense that variants of serial and parallel models can produce the same patterns of data [12]. Neural evidence has been found in support of both types of processes (see Box 1).

Like many cognitive science debates, the correct answer to the serial/parallel debate is probably "both". Consider the timing parameters of search. One can estimate the rate at which items are processed from the slopes of the reaction time (RT) by set size functions. Although the estimate depends on assumptions about factors like memory for rejected distractors (Box 2), it is in the range of 20-50 msec/item for easily identified objects that do not need to be individually fixated [26]. This estimate is significantly faster than any estimate of the total amount of time required to actually recognize an object [27]. Even on the short end, object recognition seems to require more than 100 msec/item (<10 items/second). Note that we are speaking about the time required to identify an object, not the minimum time that an observer must be exposed to an object, which can be very short, indeed [28] .

As a solution to this mismatch of times, Moore and Wolfe [20] proposed a metaphorical "carwash" (also called "pipeline" in computer science). Items might enter the binding and recognition carwash one after another every 50 msec or so. Each item might remain in the

process of recognition for several hundred milliseconds. As a consequence, if an experimenter looked at the metaphorical front or the back of the carwash, serial processing would dominate but if one looked at the carwash as a whole, one would see multiple items in the process of recognition in parallel.

Other recent models also have a serial / parallel hybrid aspect though often quite different from the carwash in detail [29,30]. Consider, for example, models of search with a primary focus on eye movements [31-33]. Here, the repeated fixations impose a form of serial selection every 250 msec or so. If one proposes that 5 or 6 items are processed in parallel at each fixation, one can produce the throughput of 20-30 items/second items found in search experiments. Interestingly, with large stimuli that can be resolved in the periphery, the pattern of response time data is similar with and without eye movements [34]. Given the close relationship of eye movements and attention [35],it could be proposed that search is accomplished by selecting successive small groups of items, whether the eyes move or not. Note that all of these versions are hybrids of some serial selection and parallel processing.

A set of basic stimulus attributes guide search

Object recognition may require attention to an object [36], but not every search requires individual scrutiny of random items before the target is attended. For example, in Figure 1, it is trivial to find the one tilted "T". Orientation is one of the basic attributes that can guide the deployment of attention. A limited set of attributes can be used to reduce the number of possible target items in a display. If you are looking for the big, red, moving vertical line, you can guide your attention toward the target size, color, motion, and orientation. We label the idea of guidance by a limited set of basic attributes as "Classic Guided Search" [37]. The set of basic attributes is not perfectly defined but there are probably between one and two dozen [38]. In the search for the green and purple Ts of Figure 1, guidance fails. T's and L's both contain a vertical and a horizontal line, so orientation information is not useful. The nature of the T or L intersection is not helpful [39], nor can guidance help by narrowing the search to the items that are both green AND purple. When you specify two features (here two colors) of the same attribute, attention is guided to the set of items that contain either purple OR green. In Figure 1, this is the set of all items [40] so no useful guidance is possible.

The internal representation of guiding attributes is different from the perceptual representation of the same attributes. What you see is not necessarily what guides your search. Consider color as an example. An item of unique color "pops out". You would have no problem finding the one red thing among yellow things [41]. The red thing looks salient and it attracts attention. It is natural to assume that the ability to guide attention is basically the same as the perceived salience of the item [42,43]. However, look for the desaturated, pale targets in Figure 2 (There are two in each panel). In each case, the target lies halfway between the saturated and white distractors in a perceptual color space. In the lab, though not in this figure, the colors can be precisely controlled so that the perceived difference between red and pale red is the same as the difference between pale green and green or pale blue and blue. Nevertheless, the desaturated red target will be found far more quickly [44], a clear dissociation between guidance and perception. Similar effects occur for other guiding attributes such as orientation [45]. The representation guiding attention should be seen as a control device, managing access to the binding and recognition bottleneck. It does not reveal itself directly in conscious perception.

Visual Search in Natural(istic) Scenes

The failure of classic guided search

To this point, this article has described what could be called "classic Guided Search" [1,37]. Now, suppose that we wanted to apply this classic Guided Search theory to the real world. Find the bread in Figure 3a. Guided Search and similar models would say that the 1-2 dozen guiding attributes define a high-dimensional space in which objects would be quite sparsely represented. That is, "bread" would be defined by some set of features [33]. If attention were guided to objects lying in the portion of the high-dimensional feature space specified by those features, few other objects would be found in the neighborhood [46]. Using a picture of the actual bread would produce better guidance than its abstract label ("bread") because more features of the specific target would be precisely described [47]. So in the real world, attention would be efficiently guided to the few bread-like objects. Guidance would massively reduce the "functional set size" [48] .

It is a good story, but it is wrong or, at least, incomplete. The story should be just as applicable to search for the loaf of bread in Figure 3b; maybe more applicable as these objects are clearly defined on a blank background. However, searches for isolated objects are quite inefficient [49] while searches like the kitchen search are very efficient (given some estimate of "set size" in real scenes) [50]. Models like Guided Search, based on bottom-up and top-down processing of a set of "preattentive" attributes seem to fail when it comes to explaining the apparent efficiency of search in the real world. Guiding attributes do some work [33, 51], but not enough.

The way forward: Expanding the concept of guidance for search in scenes

Part of the answer is that real scenes are complex, but never random. Elements are arranged in a rule-governed manner: People generally appear on horizontal surfaces [52,53], chimneys appear on roofs [54], and pots on stoves [55]. Those and other regularities of scenes can provide scene-based guidance. Borrowing from the memory literature, we will refer to "semantic" and "episodic" guidance. Semantic guidance includes knowledge of the probability of the presence an object in a scene [55] and of its probable location in that scene given the layout of the space [52,56], as well as inter-object relations (e.g. knives tend to be near forks, [57]). Violations of these expectations impede object recognition [58] and increase allocation of attention [55]. It can take longer to find a target that is semantically misplaced, (e.g., searching for the bread in the sink [59]). Episodic guidance, which we will merely mention here, refers to memory for a specific, previously encountered scene that comprises information about specific locations of specific objects [60]. Having looked several times, you know that the bread is on the counter to the left – not in all scenes, but in this one. The role of memory in search is complex (Box 2), but it is certainly the case that you will be faster, on average, to find bread in your kitchen than bread in another's kitchen.

When searching for objects in scenes, classical sources of guidance combine with episodic and semantic sources of guidance to efficiently direct our attention to those parts of the scene that have the highest probability of containing targets [52,61-63]. In naturalistic scenes, guidance of eye movements by bottom-up salience seems to play a minor role compared to guidance by more knowledge-based factors [63,64]. A short glimpse of a scene is sufficient to narrow down search space and efficiently guide gaze [65] as long as enough time is available to apply semantic knowledge to the initial scene representation [58]. However, semantic guidance cannot be too generic. Presenting a word prime (e.g. "kitchen") instead of a preview of the scene does not produce much guidance [47]. Rather, the combination of semantic scene knowledge (kitchens) with information about the structure of

the specific scene (this kitchen) seems to be crucial for effective guidance of search in real-world scenes [58,63].

A problem: Where is information about the scene coming from?

It seems reasonable to propose that semantic and episodic information about a scene guides search for objects in the scene, but where does that information come from? In order for scene information to guide attention to likely locations of "bread" in Figure 3a, you must know that the figure shows something like a kitchen. One might propose that the information about the scene develops as object after object is identified. A "kitchen" hypothesis might emerge quickly if you were lucky enough to attend first to the microwave and then to the stove, but if you were less fortunate and attended to a lamp and a window, your kitchen hypothesis might come too late to be useful.

A non-selective pathway to gist processing

Fortunately, there is another route to semantic scene information. We are able to categorize a scene as a forest without selecting individual trees for recognition [66]. A single, very brief fixation on the kitchen of Figure 3a would be enough to get the "gist" of that scene. "Gist" is an imperfectly defined term but, in this context, it includes the scene's basic-level category, an estimate of the distributions of basic attributes like color and texture [67], and the spatial layout [66,68-70]. These statistical and structural cues allow very brief exposures to support above chance categorization of scenes into, for example, natural or urban [66,71,72] or containing an animal [28,73]. Within a single fixation, Os would know that Figure 3a was a kitchen without the need to segment and identify its component objects. At 20-50 objects/second, the observer will have collected a few object identities as well but, on average, these would not be sufficient to produce categorization [66,74].

How is this possible? The answer appears to be a two-pathway architecture somewhat different from but perhaps related to previous two-pathway proposals [75,76], and somewhat different from classic two-stage, preattentive-attentive models (see Box 3). The basic idea is cartooned in Figure 4. Visual input feeds a capacity-limited "selective pathway". As described earlier, selection into the bottleneck is mediated by classic guidance and, when possible, by semantic and episodic guidance. In this two-pathway view, the raw material for semantic guidance could be generated in a non-selective pathway that is not subject to the same capacity limits. Episodic guidance would be based on the results of selective and non-selective processing.

What is a "non-selective pathway"? It is important not to invest a non-selective pathway with too many capabilities. If all processing could be done without selection and fewer capacity limits, we would not need a selective pathway. Global non-selective image processing allows observers to rapidly extract statistical information from the entire image. Observers can assess the mean and distribution of a variety of basic visual feature dimensions: size [79], orientation [80], some contrast texture descriptors [81], velocity and direction of motion [82], magnitude estimation [83], center of mass for a set of objects [84], and center of area [85]. Furthermore, summary statistics can be calculated for more complex attributes like emotion [86] or the presence of classes of objects (e.g. animal) in a scene [87].

Using these image statistics, models and (presumably) humans, can categorize scenes [66,68,69] and extract basic spatial structure [66,71]. This non-selective information could, then, provide the basis for scene-based guidance of search. Thus, non-selective categorical information, perhaps combined with the selective pathway's identification of an object or two, could strongly and rapidly suggest that Figure 3a depicts a kitchen. Non-selective structural information could give the rough layout of surfaces in the space. In principle,

these sources of information could be used to intelligently direct the resources of the selective pathway so that attention and the eyes can be deployed to likely locations of bread.

Our conscious experience of the visual world is comprised of the products of both pathways. Returning to the example at the outset of this article, when we have not yet found the object that is "right in front of our eyes", our visual experience at that location must be derived primarily from the non-selective pathway. We cannot choose to see a non-selective representation in isolation but we can gain some insight into the contributions of the two pathways from Figure 5. The non-selective pathway would 'see' the forest [66] and could provide some information about the flock of odd birds moving through it. However, identification of a tree with both green and brown boughs or of a bird heading to the right would require the work of the selective path [73].

Expert searchers like radiologists hunting for signs of cancer or airport security officers searching for threats may have learned to make specific use of non-selective signals. With some regularity, such experts will tell you that they sometimes sense the presence of a target before actually finding it. Indeed, this "Gestalt process" is a component of a leading theory of search in radiology [88]. Doctors and technicians screening for cancer can detect abnormal cases at above chance levels in a single fixation [89]. The abilities of a non-selective pathway may underpin this experience. Understanding how non-selective processing guides capacity-limited visual search could lead to improvements in search tasks that are, literally, a matter of life and death.

Concluding remarks

What is next in the study of search in scenes? We do not understand how scenes are divided up into searchable objects or proto-objects [90]. There is much work to be done to fully describe the capabilities of non-selective processing and even more to document its impact on selective processes. Finally, we would like to know if there is a neurophysiological reality to the two pathways proposed here. Suppose one "lesioned" the hypothetical selective pathway. The result might be an agnosic who could see something throughout the visual field but could not identify objects. A lesion of the non-selective pathway might produce a simultagnosic or Balint's patient, able to identify the current object of attention but otherwise unable to see. This sounds similar to the consequences of lesioning the ventral and dorsal streams, respectively [76] but more research will be required before "selective" and "non-selective" can be properly related to "what" and "where".

Box 1: Neural signatures of parallel and serial processing

What would parallel and serial processing look like at a neuronal level? One type of parallel processing in visual search is the simultaneous enhancement of all items with a preferred feature (e.g. all the red items). A number of studies have shown that for cells demonstrating a preference for a specific feature, the preference is stronger when the task is to find items with that feature [13]. For serial processing, one would like to see the "spotlight" of attention moving around from location to location. Buschman and Miller [14] saw something like this when it turned out that monkeys in their experiment liked to search a circular array of items in the same sequence on every trial. As a result, with multiple electrodes in place, they could see an attentional enhancement rise at the 3 o'clock position, then fall at 3 and rise at 6, as attention swept around in a serial manner to find a target that might be at the 9 o'clock position this time. Similar shifts of attention can be seen in human evoked potential recordings [15]. Bichot et al. [16] produced an attractive illustration of oth processes at work in visual area, V4. When the monkey was searching for "red", a cell that liked red would be more active, no matter where the monkey was looking and/or attending. If the next eye movement was going to take the

target item into the cell's receptive field, the cell showed another burst of activity as serial attention got there in advance of the eyes.

Box 2: Memory in Visual Search

There is a body of seemingly contradictory findings about the role of memory in search. First, there is the question of memory during a search. Do observers keep track of where they have been by, for example, inhibiting rejected distractors? There is some evidence for inhibition of return in visual search [17,18] though it seems clear that observers cannot use inhibition to mark every rejected distractor [19,20]. Plausibly, memory during search serves to prevent perseveration on single salient items [18,21].

What about memory for completed searches? If you find a target once, are you more efficient when you search for it again? A body of work on "repeated search" finds that search efficiency does not improve even over hundreds of trials of repetition [22,23]. On the other hand, observers can remember objects that have been seen during search [24] and implicit memory for the arbitrary layout of displays can speed response [25]. How can all of these facts be true? Of course, observers remember some results of search. (Where did I find those scissors last time?). The degree to which these memories aid subsequent search depends on whether it is faster to retrieve the relevant memory or to repeat the visual search. In many simple tasks (e.g. with arrays of letters; [23]), memory access is slower than visual search [22]. In many more commonplace searches (those scissors), memory will serve to speed the search.

Box 3: Old and new dichotomies in theories of visual search

The dichotomy between selective and non-selective pathways, proposed here, is part of a long tradition of proposing dichotomies between processes with strong capacity limits that restrict their work to one or a few objects or locations and processes that are able to operate across the entire image. It is worth briefly noting the similarities and differences with some earlier formulations.

Preattentive and Attentive processing

Preattentive processing is parallel processing over the entire image. Like non-selective processing, it is limited in its capabilities. In older formulations like Feature Integration Theory [2], it handled only basic features like color and orientation but it could be expanded to include the gist and statistical processing abilities of a non-selective pathway. The crucial difference is embodied in the term 'preattentive'. In its usual sense, preattentive processing refers to processing that occurs before the arrival in time or space of attentive processing [77]. Non-selective processing, in contrast, is proposed to occur in parallel with selective processing with the outputs of both giving rise to visual experience.

Early and Late Selection

The non-selective pathway could be seen as a form of late selection in which processing proceeds to an advanced state prior to any bottleneck in processing [78]. The selective pathway embodies early selection with only minimal processing prior to the bottleneck. Traditionally, these have been seen as competing alternatives that here coexist. Note, however, traditional late selection would permit object recognition (e.g. word recognition) prior to a bottleneck. The non-selective pathway, while able to extract some

semantic information from scenes is not proposed to have the ability to recognize objects or letters.

Acknowledgments

References

1. Wolfe JM. Guided search 2.0. A revised model of visual search. Psychon. Bull. Rev 1994;1:202–238.

2. Treisman AM, Gelade G. A feature-integration theory of attention. Cognit. Psychol 1980;12:97–136. [PubMed: 7351125]

3. Treisman A. The binding problem. Curr. Opin. Neurobiol 1996;6:171–178. [PubMed: 8725958]

4. Müller-Plath G, Elsner K. Space-based and object-based capacity limitations in visual search. Vis. Cogn 2007;15:599–634.

5. Dosher BA, et al. Information-limited parallel processing in difficult heterogeneous covert visual search. J. Exp. Psychol. Hum. Percept. Perform 2010;36:1128–1128. [PubMed: 20873936]

6. Pelli DG, Tillman KA. The uncrowded window of object recognition. Nat. Neurosci 2008;11:1129–1135. [PubMed: 18828191]

7. Balas B, et al. A summary-statistic representation in peripheral vision explains visual crowding. J. Vis 2009;9

8. Wolfe JM, Bennett SC. Preattentive object files: Shapeless bundles of basic features. Vision Res 1997;37:25–43. [PubMed: 9068829]

9. Wolfe JM. Moving towards solutions to some enduring controversies in visual search. Trends Cogn Sci 2003;7:70–76. [PubMed: 12584025]

10. Eckstein MP. The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. Psychol. Sci 1998;9:111–111.

11. Verghese P. Visual Search and Attention: A Signal Detection Theory Approach. Neuron 2001;31:523–535. [PubMed: 11545712]

12. Townsend JT, Wenger MJ. The serial-parallel dilemma: A case study in a linkage of theory and method. Psychon. Bull. Rev 2004;11:391–391. [PubMed: 15376788]

13. Chelazzi L, et al. A neural basis for visual search in inferior temporal cortex. Nature 1993;363:345–347. [PubMed: 8497317]

14. Buschman TJ, Miller EK. Serial, covert shifts of attention during visual search are reflected by the frontal eye fields and correlated with population oscillations. Neuron 2009;63:386–396. [PubMed: 19679077]

15. Woodman GF, Luck SJ. Serial deployment of attention during visual search. J. Exp. Psychol. Hum. Percept. Perform 2003;29:121–138. [PubMed: 12669752]

16. Bichot NP, et al. Parallel and serial neural mechanisms for visual search in macaque area V4. Science 2005;308:529–529. [PubMed: 15845848]

17. Takeda Y, Yagi A. Inhibitory tagging in visual search can be found if search stimuli remain visible. Percept. Psychophys 2000;62:927–934. [PubMed: 10997039]

18. Klein R. On the Control of Attention. Canadian Journal of Experimental Psychology 2009;63:240–252. [PubMed: 19739907]

19. Horowitz TS, Wolfe JM. Visual search has no memory. Nature 1998;394:575–577. [PubMed: 9707117]

20. Moore, CM.; Wolfe, JM. Getting beyond the serial/parallel debate in visual search: A hybrid approach. In: Shapiro, K., editor. The limits of attention: Temporal constraints on human information processing. Oxford U. Press; 2001. p. 178-198.

21. Klein RM, MacInnes WJ. Inhibition of return is a foraging facilitator in visual search. Psychol. Sci 1999;10:346–346.

22. Kunar MA, et al. The role of memory and restricted context in repeated visual search. Percept. Psychophys 2008;70:314–328. [PubMed: 18372752]

23. Wolfe JM, et al. Postattentive vision. J. Exp. Psychol. Hum. Percept. Perform 2000;26:693–716. [PubMed: 10811170]

24. Hollingworth A, Henderson JM. Accurate visual memory for previously attended objects in natural scenes. J. Exp. Psychol. Hum. Percept. Perform 2002;28:113–136.

25. Jiang Y, Wagner LC. What is learned in spatial contextual cuing-configuration or individual locations? Percept. Psychophys 2004;66:454–463. [PubMed: 15283070]

26. Horowitz TS. Revisiting the variable memory model of visual search. Vis. Cogn 2006;14:668–684.

27. Theeuwes J, et al. A new estimation of the duration of attentional dwell time. Psychon. Bull. Rev 2004;11:60–60. [PubMed: 15116987]

28. Kirchner H, Thorpe SJ. Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. Vision Res 2006;46:1762–1776. [PubMed: 16289663]

29. Thornton TL, Gilden DL. Parallel and serial processes in visual search. Psychol. Rev 2007;114:71–103. [PubMed: 17227182]

30. Fazl A, et al. View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds. Cognit. Psychol 2009;58:1–48. [PubMed: 18653176]

31. Renninger LW, et al. Where to look next? Eye movements reduce local uncertainty. J. Vis 2007;7

32. Geisler WS, et al. Visual search: The role of peripheral information measured using gaze-contingent displays. J. Vis 2006;6:858–873. [PubMed: 17083280]

33. Zelinsky GJ. A theory of eye movements during target acquisition. Psychol. Rev 2008;115:787–787. [PubMed: 18954205]

34. Zelinsky GJ, Sheinberg DL. Eye movements during parallel-serial visual search. J. Exp. Psychol. Hum. Percept. Perform 1997;23:244–244. [PubMed: 9090154]

35. Kowler E, et al. The role of attention in the programming of saccades. Vision Res 1995;35:1897–1916. [PubMed: 7660596]

36. Huang L. What Is the Unit of Visual Attention? Object for Selection, but Boolean Map for Access. J. Exp. Psychol. Gen 2010;139:162–179. [PubMed: 20121317]

37. Wolfe, JM. Guided Search 4.0: Current Progress with a model of visual search. In: Gray, W., editor. Integrated models of cognitive systems. Oxford; New York: 2007. p. 99-119.

38. 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]

39. Wolfe JM, DiMase JS. Do intersections serve as basic features in visual search? Percept 2003;32:645–656.

40. Wolfe JM, et al. Limitations on the Parallel Guidance of Visual Search: Color x Color and Orientation x Orientation Conjuctions. J. Exp. Psychol. Hum. Percept. Perform 1990;16:879–892. [PubMed: 2148599]

41. Treisman A, Gormican S. Feature analysis in early vision: Evidence from search asymmetries. Psychol. Rev 1988;95:15–48. [PubMed: 3353475]

42. Parkhurst D, et al. Modeling the role of salience in the allocation of overt visual attention. Vision Res 2002;42:107–123. [PubMed: 11804636]

43. Itti L, et al. A model of saliency-based visual attention for rapid scene analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on 2002;20:1254–1259.

44. Lindsey DT, et al. Color channels, not color appearance or color categories, guide visual search for desaturated color targets. Psychol. Sci 2010;21:1208–1208. [PubMed: 20713637]

45. Wolfe JM, et al. The role of categorization in visual search for orientation. J. Exp. Psychol. Hum. Percept. Perform 1992;18:34–49. [PubMed: 1532193]

46. DiCarlo JJ, Cox DD. Untangling invariant object recognition. Trends Cogni. Sci 2007;11:333–341.

47. Castelhano MS, Heaven C. The relative contribution of scene context and target features to visual search in scenes. Atten. Percept. Psychophys 2010;72:1283–1297. [PubMed: 20601710]

48. Neider MB, Zelinsky GJ. Exploring set size effects in scenes: Identifying the objects of search. Vis. Cogn 2008;16:1–10.

49. Vickery TJ, et al. Setting up the target template in visual search. J. Vis 2005;5:81–92. [PubMed: 15831069]

50. Wolfe J, et al. Search for arbitrary objects in natural scenes is remarkably efficient. J. Vis 2008;8:1103–1103.

51. Kanan C, et al. SUN: Top-down saliency using natural statistics. Vis. Cogn 2009;17:979–1003. [PubMed: 21052485]

52. Torralba A, et al. Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. Psychol. Rev 2006;113:766–786. [PubMed: 17014302]

53. Droll J, Eckstein M. Expected object position of two hundred fifty observers predicts first fixations of seventy seven separate observers during search. J. Vis 2008;8:320–320.

54. Eckstein MP, et al. Attentional cues in real scenes, saccadic targeting, and Bayesian priors. Psychol. Sci 2006;17:973–980. [PubMed: 17176430]

55. Võ MLH, Henderson JM. Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. J. Vis 2009;9:1–15.

56. Võ MLH, Henderson JM. The time course of initial scene processing for eye movement guidance in natural scene search. J. Vis 2010;10:1–13.

57. Bar M. Visual objects in context. Nat. Rev. Neurosci 2004;5:617–629. [PubMed: 15263892]

58. Biederman I, et al. Scene perception: Detecting and judging objects undergoing relational violations. Cognit. Psychol 1982;14:143–177. [PubMed: 7083801]

59. Malcolm GL, Henderson JM. Combining top-down processes to guide eye movements during real-world scene search. J. Vis 2009;10:1–11.

60. Hollingworth A. Scene and position specificity in visual memory for objects. J.Exp. Psychol. Learn.Mem.Cogn 2006;32:58–69. [PubMed: 16478340]

61. Neider MB, Zelinsky GJ. Scene context guides eye movements during visual search. Vision Res 2006;46:614–621. [PubMed: 16236336]

62. Ehinger KA, et al. Modelling search for people in 900 scenes: A combined source model of eye guidance. Vis. Cogn 2009;17:945–978. [PubMed: 20011676]

63. Henderson, et al. Searching in the dark: Cognitive relevance drives attention in real-world scenes. Psychon. Bull. Rev 2009;16:850–856. [PubMed: 19815788]

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

65. Castelhano MS, Henderson JM. Initial scene representations facilitate eye movement guidance in visual search. J. Exp. Psychol. Hum. Percept. Perform 2007;33:753–763. [PubMed: 17683226]

66. Greene MR, Oliva A. Recognition of natural scenes from global properties: Seeing the forest without representing the trees. Cognit. Psychol 2009;58:137–176. [PubMed: 18762289]

67. Rousselet G, et al. How long to get to the "gist" of real-world natural scenes? Vis. Cogn 2005;12:852–877.

68. Sanocki T. Representation and perception of scenic layout. Cognit. Psychol 2003;47:43–86. [PubMed: 12852935]

69. Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope. Int. J. Comput. Vis 2001;42:145–175.

70. Biederman I, et al. On the information extracted from a glance at a scene. J. Exp. Psychol. Hum. Percept. Perform 1974;103:597–600.

71. Greene MR, Oliva A. The Briefest of Glances: The Time Course of Natural Scene Understanding. Psychol. Sci 2009;20:464–472. [PubMed: 19399976]

72. Joubert OR, et al. Processing scene context: Fast categorization and object interference. Vision Res 2007;47:3286–3297. [PubMed: 17967472]

73. Evans KK, Treisman A. Perception of objects in natural scenes: is it really attention free? J. Exp. Psychol. Hum. Percept. Perform 2005;31:1476–1492. [PubMed: 16366803]

74. Joubert OR, et al. Early interference of context congruence on object processing in rapid visual categorization of natural scenes. J. Vis 2008;8:11–11. [PubMed: 19146341]

75. Held, R. The neurosciences: Second study program. Rockefeller U.; New York: 1970. Two modes of processing spatially distributed visual stimulation; p. 317-324.

76. Ungerleider, LG.; Mishkin, M. Analysis of visual behavior. MIT Press; 1982. Two cortical visual systems; p. 586-586.

77. Neisser, U. Cognitive psychology. Appleton-Century-Crofts; 1967.

78. Deutsch JA, Deutsch D. Attention: Some theoretical considerations. Psychol. Rev 1963;70:80–90. [PubMed: 14027390]

79. Chong SC, Treisman A. Representation of statistical properties. Vision Res 2003;43:393–404. [PubMed: 12535996]

80. Parkes L, et al. Compulsory averaging of crowded orientation signals in human vision. Nat. Neurosci 2001;4:739–744. [PubMed: 11426231]

81. Chubb C, et al. The three dimensions of human visual sensitivity to first-order contrast statistics. Vision Res 2007;47:2237–2248. [PubMed: 17619044]

82. Williams DW, Sekuler R. Coherent global motion percepts from stochastic local motions. Vision Res 1984;24:55–62. [PubMed: 6695508]

83. Demeyere N, et al. Automatic statistical processing of visual properties in simultanagnosia. Neuropsychologia 2008;46:2861–2864. [PubMed: 18589460]

84. Alvarez GA, Oliva A. The representation of simple ensemble visual features outside the focus of attention. Psychol. Sci 2008;19:392–398. [PubMed: 18399893]

85. Melcher D, Kowler E. Shapes, surfaces and saccades. Vision Res 1999;39:2929–2946. [PubMed: 10492819]

86. Haberman J, Whitney D. Rapid extraction of mean emotion and gender from sets of faces. Curr. Biol 2007;17:R751–R753. [PubMed: 17803921]

87. Vanrullen R. Binding hardwired versus on-demand feature conjunctions. Vis. Cogn 2009;17:103–119.

88. Krupinski EA. Current perspectives in medical image perception. Atten. Percept. Psychophys 2010;72:1205–1217. [PubMed: 20601701]

89. Kundel HL, Nodine CF. Interpreting chest radiographs without visual search. Radiology 1975;116:527–532. [PubMed: 125436]

90. Rensink RA. The dynamic representation of scenes. Vis. Cogn 2000;7:17–42.
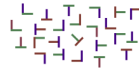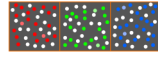
Figure 1.
Find the four purple and green Ts. Even though it is easy to identify such targets, this task requires search.

Figure 2.
Find the desaturated color dots. Colors are only an approximation of the colors that would be used in a carefully calibrated experiment. The empirical result is that it is much easier to find the pale red (pink) targets than to find pale green or blue.

Figure 3.
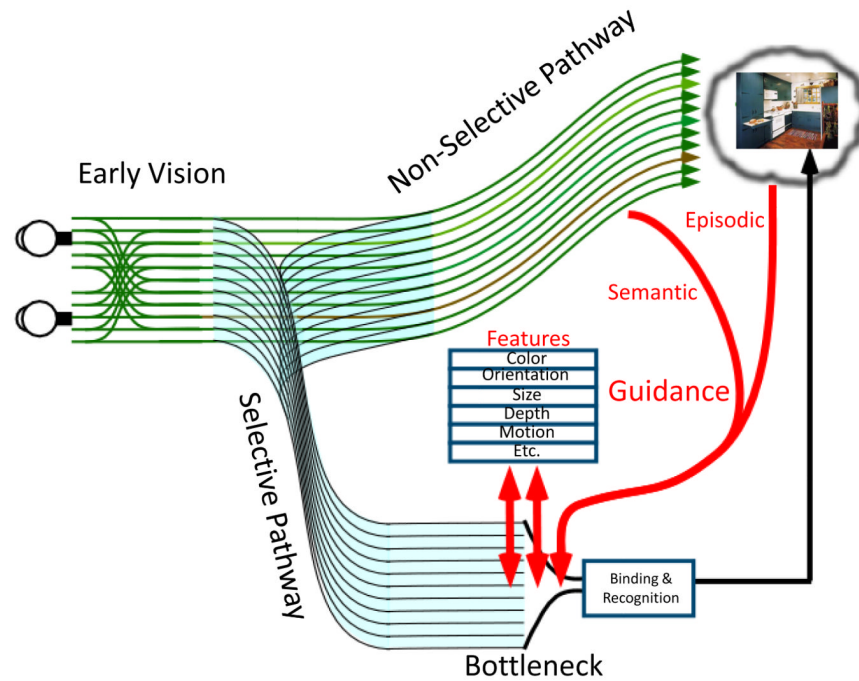Find the loaf of bread in each panel.

Figure 4.
A two-pathway architecture for visual processing. A selective pathway can bind features and recognize objects, but it is severely capacity-limited. The limit is shown as a "bottleneck" in the pathway. Access to the bottleneck is controlled by guidance mechanisms that allow items that are more likely to be targets preferential access to feature binding and object recognition. Classic guidance, cartooned in the box above the bottleneck, gives preference to items with basic target features (e.g. color). This paper posits scene guidance (semantic and episodic), with semantic guidance derived from a non-selective pathway. This non-selective pathway can extract statistics from the entire scene, enabling a certain amount of semantic processing but not precise object recognition.
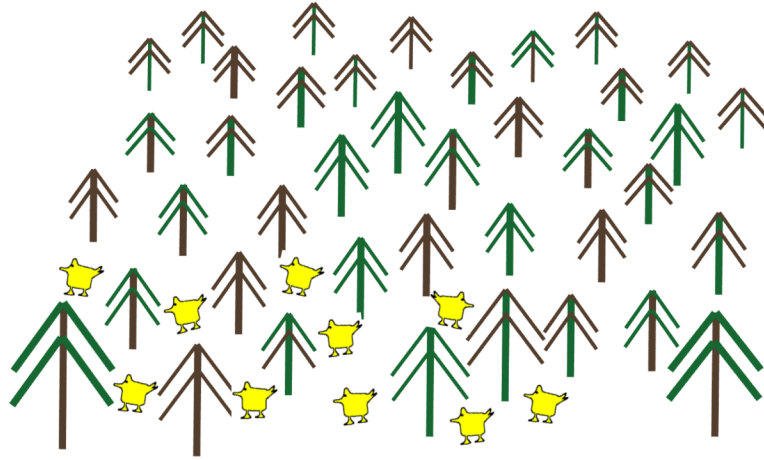
Figure 5.
What do you see? And how does that change when you are asked to look for an untilted bird or trees with brown trunks and green boughs? It is proposed that a non-selective pathway would 'see' image statistics like average color or orientation in a region. It could get the 'gist' of forest and, perhaps, the presence of animals. It would not know which trees had brown trunks or which birds were tilted.