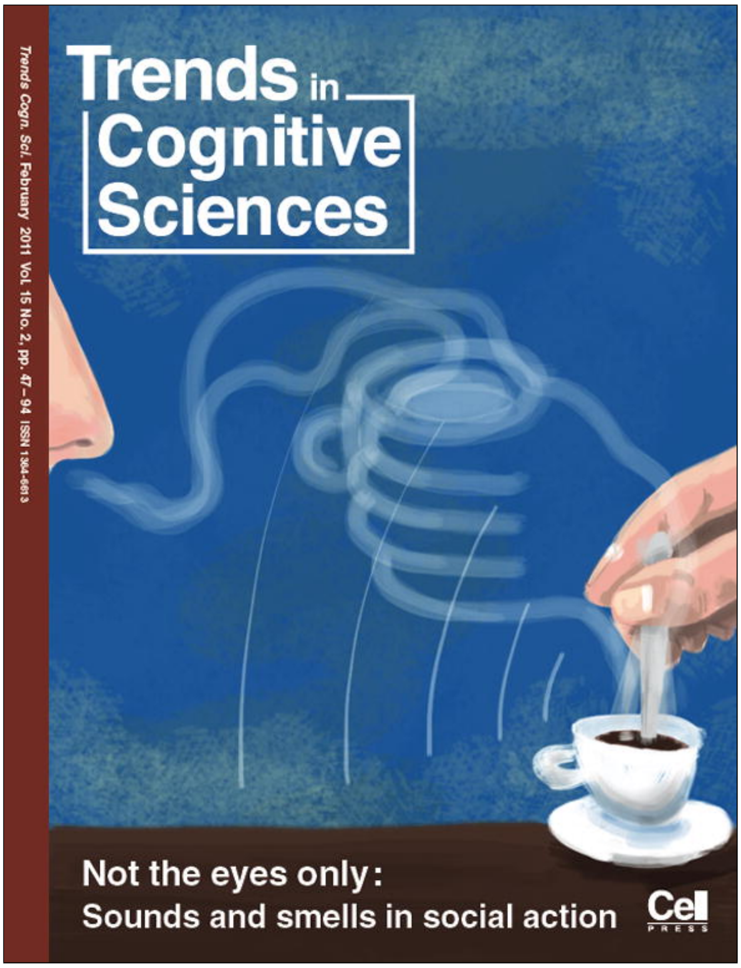
Trends in Cognitive Sciences
Trends Cogn. Sci. February 2011 Vol. 15 No. 2, pp. 47–94 ISSN 1364-6613
Not the eyes only:
Sounds and smells in social action
Cell PRESS

Review

Visual search in scenes involves selective and nonselective pathways

Jeremy M. Wolfe, Melissa L.-H. Võ, Karla K. Evans and Michelle R. Greene

Brigham & Women's Hospital, Harvard Medical School, 64 Sidney St. Suite 170, Cambridge, MA 02139, USA

How does one 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 article 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 might be best explained by a dual-path model: a 'selective' path in which candidate objects must be individually selected for recognition and a 'nonselective' path in which information can be extracted from global and/or statistical information.

Searching and experiencing a scene

It is an interesting aspect of visual experience that you can look for an object that is, literally, right in front of your eyes, yet not find it for an appreciable period of time. It is clear that you are seeing something at the location of the object before you find it. What is that something and how do you go about finding that desired object? These questions have occupied visual search researchers for decades. Whereas 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. During the past 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. In this article, we 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 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 tells you 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 nonselective processing is put to use when you 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 we briefly review below, it holds that search is necessary because object recognition processes are limited to one or, perhaps, a 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, such as color, orientation and size.

Object recognition is capacity limited

You need to search because, although you are good at recognizing objects, you cannot recognize multiple objects simultaneously. For example, all of the objects in Figure 1 are simple in construction, but if you are asked to find 'T's that are both 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, 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, although it would be a less compelling demonstration, it would still be necessary to attend to item after item to bind their features and recognize them even if there were only a few items and even if those were widely spaced [8].

The selection mechanism is a serial–parallel hybrid

Whereas 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 simultaneously 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, are 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 process (Box 1).

Similar to many cognitive science debates, the correct answer to the serial–parallel debate is probably 'both'. Consider the timing parameters of search. One can esti-

Corresponding author: Wolfe, J.M. (wolfe@search.bwh.harvard.edu).

1364-6613/$ – see front matter doi:10.1016/j.tics.2010.12.001 Trends in Cognitive Sciences, February 2011, Vol. 15, No. 2

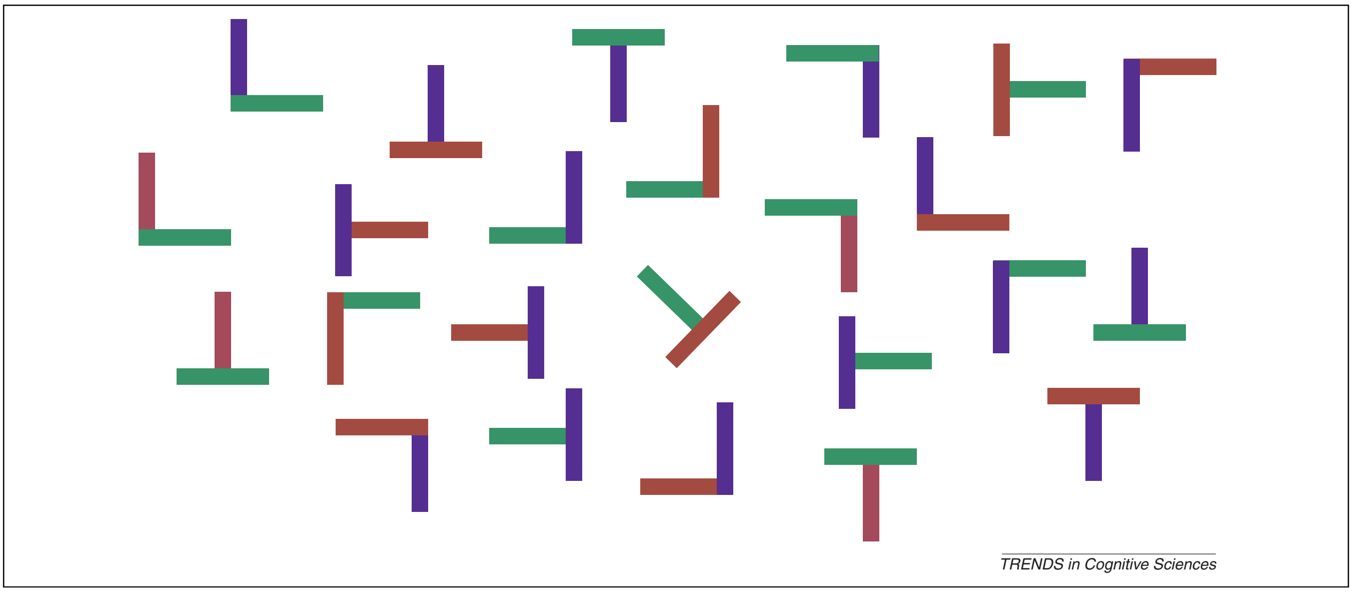

Figure 1. Find the four purple-and-green Ts. Even though it is easy to identify such targets, this task requires search.

mate 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 such as 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 [13]. This estimate is significantly faster than any estimate of the total amount of time required to recognize an object [14]. Even on the short end, object recognition seems to require more than 100 msec/item (<10 items/sec). 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 [15].

As a solution to this mismatch of times, Moore and Wolfe [16] 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, although they are often different from the carwash in detail [17,18]. Consider, for example, models of search with a primary focus on eye movements [19–21]. Here, the repeated fixations impose a form of serial selection every 250 msec or so. If one proposes that five or six items are processed in parallel at each fixation, one can produce the throughput of 20–30 items/second found in search experiments. Interestingly, with large stimuli that can be re-

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). Several 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 [77]. For serial processing, one would like to see the 'spotlight' of attention moving around from location to location. Buschman and Miller [78] saw something similar to 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, the authors 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 in that particular trial.

Similar shifts of attention can be seen in human evoked potential recordings [79]. Bichot *et al.* [80] produced an attractive illustration of both 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 receptive field of the cell, the cell showed another burst of activity as serial attention reached it 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, for example, by inhibiting rejected distractors? There is some evidence for inhibition of return in visual search [81,82], although it seems clear that observers cannot use inhibition to mark every rejected distractor [16,83]. Plausibly, memory during search serves to prevent perseveration on single salient items [82,84].

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 [85,86]. By contrast, observers can remember objects that have been seen during search [87] and implicit memory for the arbitrary layout of displays can speed their response [88]. 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; [86]), memory access is slower than is visual search [85]. In many more commonplace searches (those scissors), memory will serve to speed the search.

solved in the periphery, the pattern of response time data is similar with and without eye movements [22]. Given the close relationship of eye movements and attention [23], 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 might require attention to an object [24], 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' [25]. The set of basic attributes is not perfectly defined but there are probably between one and two dozen [26]. In the search for the green-and-purple Ts of Figure 1, guidance fails. Ts and Ls both contain a vertical and a horizontal line, so orientation information is not useful. The nature of the T or L intersection is also not helpful [27]; neither 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 [28] 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 [29]. 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 [30,31]. 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, although 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 more quickly [32], a clear dissociation between guidance and perception. Similar effects occur for other guiding attributes, such as orientation [33]. 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, we have described what could be called 'classic guided search' [1,25]. 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 one to two 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 [21]. 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 [34]. 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 [35]. So in the real world, attention would be efficiently guided to the few bread-like objects. Guidance would reduce the 'functional set size' [36].

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 inefficient [37], whereas searches such as the kitchen search are efficient (given some estimate of 'set size' in real scenes) [38]. Models such as 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 [21,39], but not enough.

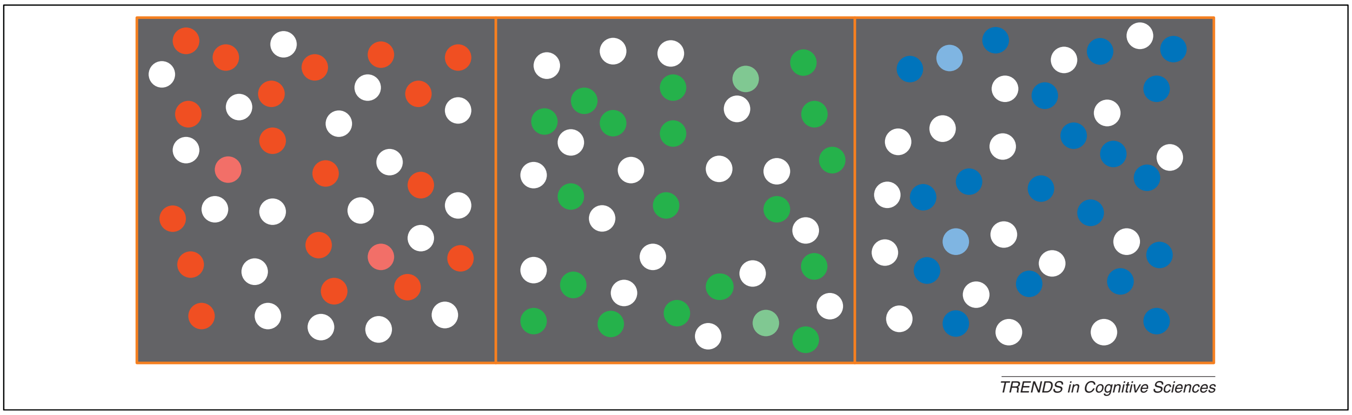

Figure 2. Find the desaturated color dots. Colors are only an approximation of the colors that would be used in a carefully calibrated experiment. The empirical result is that it is easier to find the pale-red (pink) targets than to find the pale-green or -blue targets.

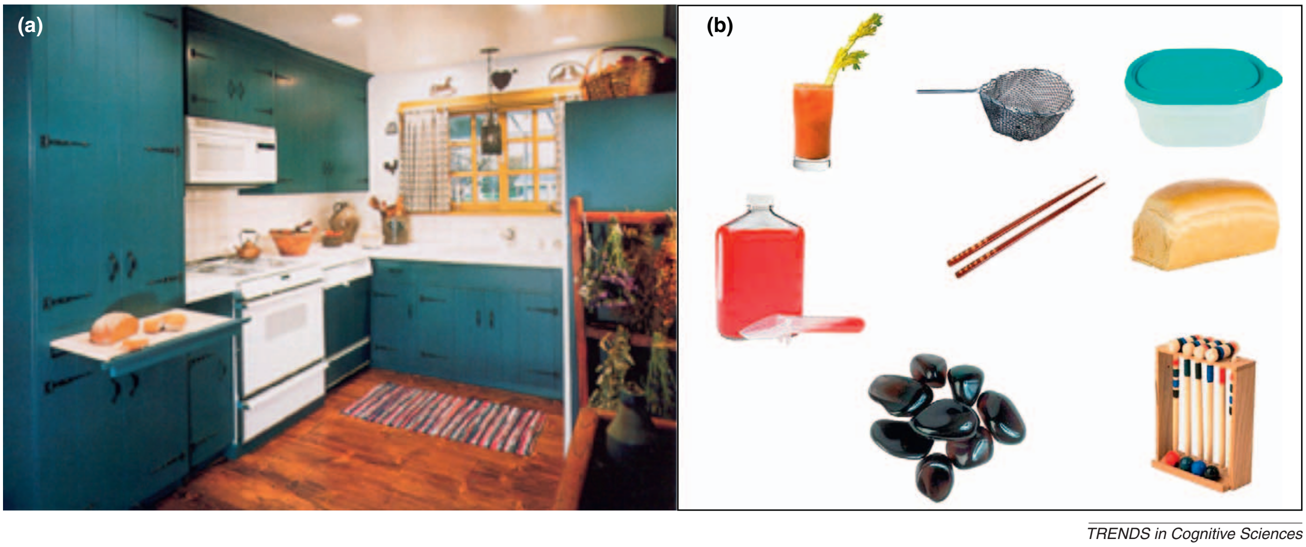

Figure 3. Find the loaf of bread in each of **(a)** and **(b)**.

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 [40,41], chimneys appear on roofs [42] and pots on stoves [43]. Those and other regularities of scenes can provide scene-based guidance. Borrowing from the memory literature, we refer to 'semantic' and 'episodic' guidance. Semantic guidance includes knowledge of the probability of the presence of an object in a scene [43] and of its probable location in that scene given the layout of the space [40,44], as well as inter-object relations (e.g. knives tend to be near forks, [45]). Violations of these expectations impede object recognition [46] and increase allocation of attention [43]. It can take longer to find a target that is semantically misplaced, (e.g. searching for the bread in the sink [47]). 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 [48]. 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 the case that you will be faster, on average, to find bread in your kitchen than in another's kitchen.

When searching for objects in scenes, classic sources of guidance combine with episodic and semantic sources of guidance to direct your attention efficiently to those parts of the scene that have the highest probability of containing targets [40,49–51]. In naturalistic scenes, guidance of eye movements by bottom-up salience seems to play a minor role compared with guidance by more knowledge-based factors [51,52]. A short glimpse of a scene is sufficient to narrow down search space and efficiently guide gaze [53] as long as enough time is available to apply semantic knowledge to the initial scene representation [44]. 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 [35]. 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 [44,51].

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? For scene information to guide attention to probable 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 nonselective pathway to gist processing

Fortunately, there is another route to semantic scene information. Humans are able to categorize a scene as a forest without selecting individual trees for recognition [54]. A single, 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 basic-level category of the scene, an estimate of the distributions of basic attributes, such as color and texture [55], and the spatial layout [54,56–58]. These statistical and structural cues allow brief exposures to support above-chance categorization of scenes into, for example, natural or urban [54,59,60] or containing an animal [15,61]. Within a single fixation, an observer would know that Figure 3a was a kitchen without the need to segment and identify its component objects. At 20–50 objects/second, that observer will have collected a few object identities as well but, on average, these would not be sufficient to produce categorization [54,62].

How is this possible? The answer appears to be a two-pathway architecture somewhat different from, but perhaps related to, previous two-pathway proposals [63,64], and somewhat different from classic two-stage, preattentive-attentive models (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 nonselective pathway that is not subject to the same capacity limits. Episodic guidance would be based on the results of selective and nonselective processing.

What is a 'nonselective pathway'? It is important not to invest a nonselective pathway with too many capabilities. If all processing could be done without selection and fewer capacity limits, one would not need a selective pathway. Global nonselective image processing allows observers to extract statistical information rapidly from the entire image. Observers can assess the mean and distribution of a variety of basic visual feature dimensions: size [65], orientation [66], some contrast texture descriptors [67], velocity and direction of motion [68], magnitude estimation [69], center of mass for a set of objects [70] and center of area [71]. Furthermore, summary statistics can be calculated for more complex attributes, such as emotion [72] or the presence of classes of objects (e.g. animal) in a scene [73].

Using these image statistics, models and (presumably) humans, can categorize scenes [54,56,57] and extract basic

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

The dichotomy between selective and nonselective 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. Similar to nonselective processing, it is limited in its capabilities. In older formulations such as Feature Integration Theory [2], it handled only basic features, such as color and orientation, but it could be expanded to include the gist and statistical-processing abilities of a nonselective 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 [89]. Nonselective processing, by 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 nonselective pathway could be seen as a form of late selection in which processing proceeds to an advanced state before any bottleneck in processing [90]. The selective pathway embodies early selection with only minimal processing before the bottleneck. Traditionally, these have been seen as competing alternatives that coexist here. However, traditional late selection would permit object recognition (e.g. word recognition) before a bottleneck. The nonselective pathway, although able to extract some semantic information from scenes, is not proposed to have the ability to recognize either objects or letters.

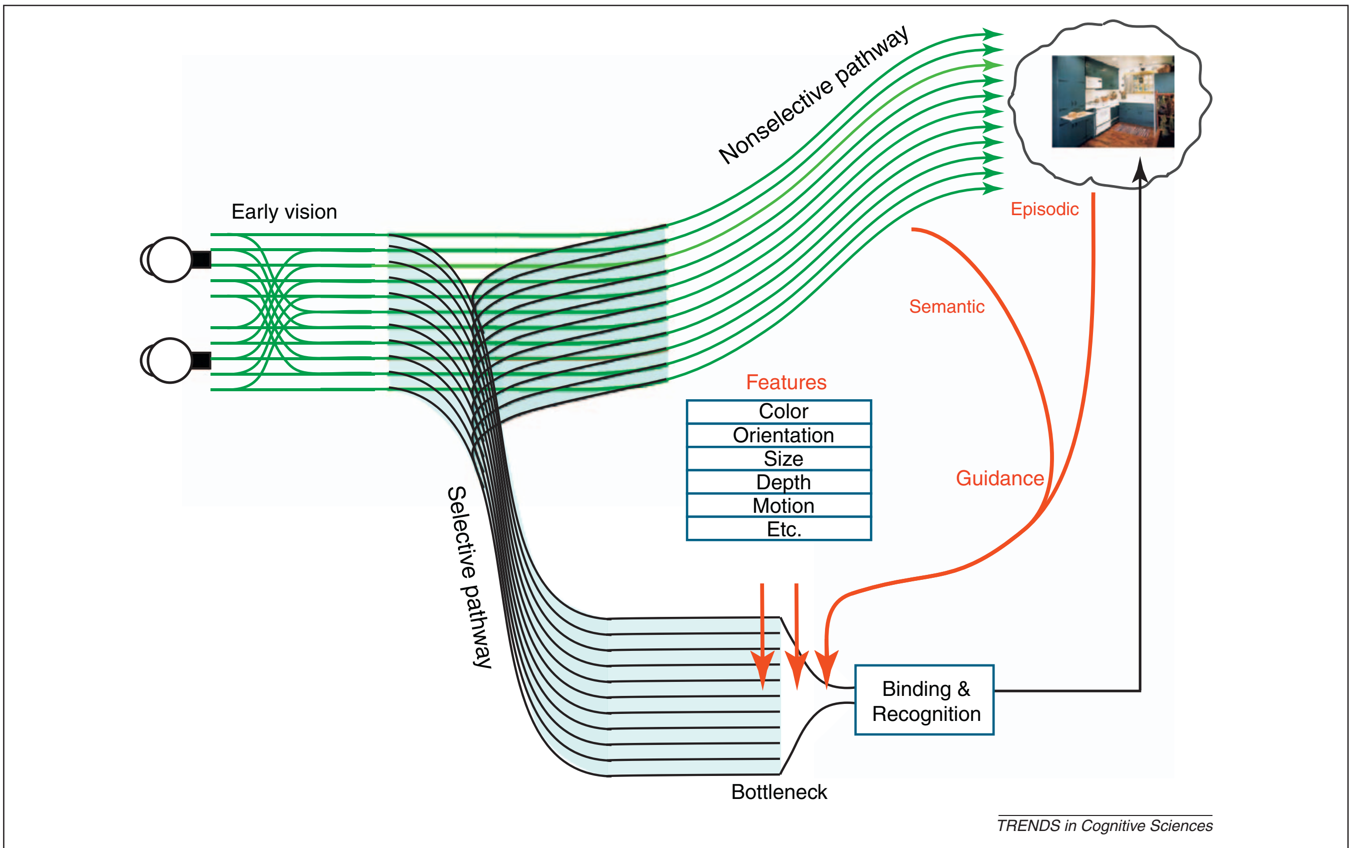

Figure 4. A two-pathway architecture for visual processing. A selective pathway can bind features and recognize objects, but it is capacity limited. The limit is shown as a 'bottleneck' in the pathway. Access to the bottleneck is controlled by guidance mechanisms that allow items that are more likely 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 article posits scene guidance (semantic and episodic), with semantic guidance derived from a nonselective pathway. This nonselective pathway can extract statistics from the entire scene, enabling a certain amount of semantic processing, but not precise object recognition.

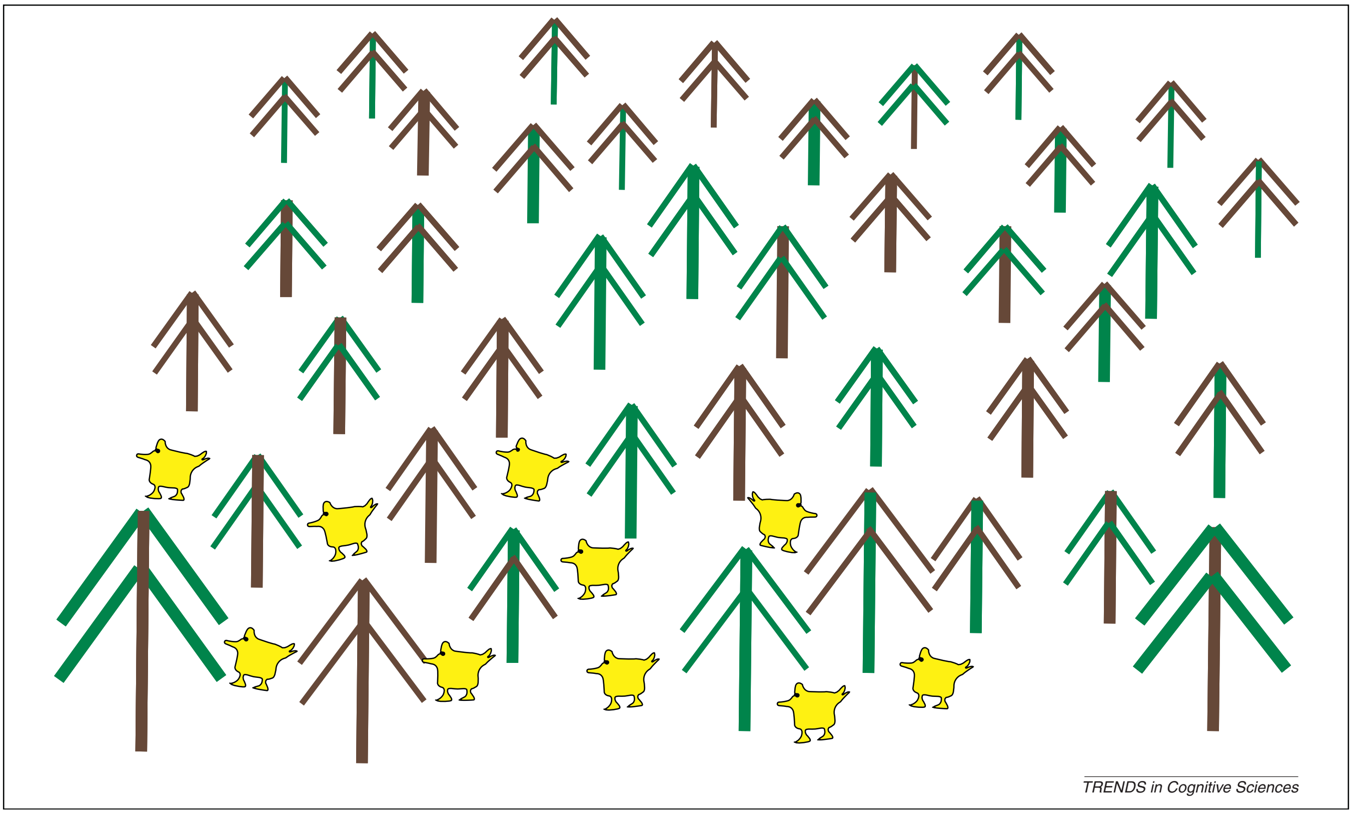

Figure 5. What do you see? How does that change when you are asked to look for an untilted bird or trees with brown trunks and green boughs? It is proposed that a nonselective pathway would 'see' image statistics, such as average color or orientation, in a region. It could get the 'gist' of forest and, perhaps, the presence of animals. However, it would not know which trees had brown trunks or which birds were tilted.

spatial structure [54,59]. This nonselective information could then provide the basis for scene-based guidance of search. Thus, nonselective categorical information, perhaps combined with the identification of an object or two by the selective pathway, could strongly and rapidly suggest that Figure 3a depicts a kitchen. Nonselective structural information could give the rough layout of surfaces in the space. In principle, these sources of information could be used to direct the resources of the selective pathway intelligently so that attention and the eyes can be deployed to probable locations of bread.

Your 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 you have not yet found the object that is 'right in front of your eyes', your visual experience at that location must be derived primarily from the nonselective pathway. You cannot choose to see a nonselective representation in isolation, but you can gain some insight into the contributions of the two pathways from Figure 5. The nonselective pathway would 'see' the forest [54] 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 [61].

Expert searchers, such as radiologists hunting for signs of cancer or airport security officers searching for threats, might have learned to make specific use of nonselective signals. With some regularity, such experts will tell you that they sometimes sense the presence of a target before finding it. Indeed, this 'Gestalt process' is a component of a leading theory of search in radiology [74]. Doctors and technicians screening for cancer can detect abnormal cases at above-chance levels in a single fixation [75]. The abilities of a nonselective pathway might underpin this experience. Understanding how nonselective 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? It is still not understood how scenes are divided up into searchable objects or proto-objects [76]. There is much work to be done to describe fully the capabilities of nonselective 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 nonselective 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 [64], but more research will be required before 'selective' and 'nonselective' can be properly related to 'what' and 'where'.

Acknowledgments

This work was supported by NIH EY017001 and ONR MURI N000141010278 to J.M.W. K.K.E. was supported by NIH/NEI 1F32EY019819-01, M.R.G. by NIH/NEI F32EY019815-01and M.L-H.V. by DFG 1683/1-1.

References

1 Wolfe, J.M. (1994) Guided search 2.0. A revised model of visual search. *Psychon. Bull. Rev.* 1, 202–238
2 Treisman, A.M. and Gelade, G. (1980) A feature-integration theory of attention. *Cognit. Psychol.* 12, 97–136
3 Treisman, A. (1996) The binding problem. *Curr. Opin. Neurobiol.* 6, 171–178
4 Müller-Plath, G. and Elsner, K. (2007) Space-based and object-based capacity limitations in visual search. *Vis. Cogn.* 15, 599–634
5 Dosher, B.A. *et al.* (2010) Information-limited parallel processing in difficult heterogeneous covert visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1128–11128
6 Pelli, D.G. and Tillman, K.A. (2008) The uncrowded window of object recognition. *Nat. Neurosci.* 11, 1129–1135
7 Balas, B. *et al.* (2009) A summary-statistic representation in peripheral vision explains visual crowding. *J. Vis.* 9, 1–18
8 Wolfe, J.M. and Bennett, S.C. (1997) Preattentive object files: shapeless bundles of basic features. *Vision Res.* 37, 25–43
9 Wolfe, J.M. (2003) Moving towards solutions to some enduring controversies in visual search. *Trends Cogn. Sci.* 7, 70–76
10 Eckstein, M.P. (1998) The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychol. Sci.* 9, 111–1111
11 Verghese, P. (2001) Visual search and attention: a signal detection theory approach. *Neuron* 31, 523–535
12 Townsend, J.T. and Wenger, M.J. (2004) The serial–parallel dilemma: a case study in a linkage of theory and method. *Psychon. Bull. Rev.* 11, 391–1391
13 Horowitz, T.S. (2006) Revisiting the variable memory model of visual search. *Vis. Cogn.* 14, 668–684
14 Theeuwes, J. *et al.* (2004) A new estimation of the duration of attentional dwell time. *Psychon. Bull. Rev.* 11, 60–160
15 Kirchner, H. and Thorpe, S.J. (2006) Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Res.* 46, 1762–1776
16 Moore, C.M. and Wolfe, J.M. (2001) Getting beyond the serial/parallel debate in visual search: a hybrid approach. In *The Limits of Attention: Temporal Constraints on Human Information Processing* (Shapiro, K., ed.), pp. 178–198, Oxford University Press
17 Thornton, T.L. and Gilden, D.L. (2007) Parallel and serial processes in visual search. *Psychol. Rev.* 114, 71–103
18 Fazl, A. *et al.* (2009) View-invariant object category learning, recognition, and search: how spatial and object attention are coordinated using surface-based attentional shrouds. *Cognit. Psychol.* 58, 1–48
19 Renninger, L.W. *et al.* (2007) Where to look next? Eye movements reduce local uncertainty. *J. Vis.* 7, 1–17
20 Geisler, W.S. *et al.* (2006) Visual search: the role of peripheral information measured using gaze-contingent displays. *J. Vis.* 6, 858–873
21 Zelinsky, G.J. (2008) A theory of eye movements during target acquisition. *Psychol. Rev.* 115, 787–1787
22 Zelinsky, G.J. and Sheinberg, D.L. (1997) Eye movements during parallel-serial visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 244–1244
23 Kowler, E. *et al.* (1995) The role of attention in the programming of saccades. *Vision Res.* 35, 1897–1916
24 Huang, L. (2010) What is the unit of visual attention? Object for selection, but Boolean map for access. *J. Exp. Psychol. Gen.* 139, 162–179
25 Wolfe, J.M. (2007) Guided Search 4.0: current Progress with a model of visual search. In *Integrated Models of Cognitive Systems* (Gray, W., ed.), pp. 99–119, Oxford University Press
26 Wolfe, J.M. and Horowitz, T.S. (2004) What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.* 5, 495–501
27 Wolfe, J.M. and DiMase, J.S. (2003) Do intersections serve as basic features in visual search? *Perception* 32, 645–656
28 Wolfe, J.M. *et al.* (1990) Limitations on the parallel guidance of visual search: color × color and orientation × orientation conjunctions. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 879–892
29 Treisman, A. and Gormican, S. (1988) Feature analysis in early vision: evidence from search asymmetries. *Psychol. Rev.* 95, 15–48
30 Parkhurst, D. *et al.* (2002) Modeling the role of salience in the allocation of overt visual attention. *Vision Res.* 42, 107–123
31 Itti, L. *et al.* (2002) A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Machine Intell.* 20, 1254–1259
32 Lindsey, D.T. *et al.* (2010) Color channels, not color appearance or color categories, guide visual search for desaturated color targets. *Psychol. Sci.* 21, 1208–11208
33 Wolfe, J.M. *et al.* (1992) The role of categorization in visual search for orientation. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 34–49
34 DiCarlo, J.J. and Cox, D.D. (2007) Untangling invariant object recognition. *Trends Cogn. Sci.* 11, 333–341
35 Castelhano, M.S. and Heaven, C. (2010) The relative contribution of scene context and target features to visual search in scenes. *Atten. Percept. Psychophys.* 72, 1283–1297
36 Neider, M.B. and Zelinsky, G.J. (2008) Exploring set size effects in scenes: identifying the objects of search. *Vis. Cogn.* 16, 1–10
37 Vickery, T.J. *et al.* (2005) Setting up the target template in visual search. *J. Vis.* 5, 81–92
38 Wolfe, J. *et al.* (2008) Search for arbitrary objects in natural scenes is remarkably efficient. *J. Vis.* 8, 1103–11103
39 Kanan, C. *et al.* (2009) SUN: top-down saliency using natural statistics. *Vis. Cogn.* 17, 979–1003
40 Torralba, A. *et al.* (2006) Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol. Rev.* 113, 766–786
41 Droll, J. and Eckstein, M. (2008) Expected object position of two hundred fifty observers predicts first fixations of seventy seven separate observers during search. *J. Vis.* 8, 320–1320
42 Eckstein, M.P. *et al.* (2006) Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychol. Sci.* 17, 973–980
43 Võ, M.L.-H. and Henderson, J.M. (2009) Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *J. Vis.* 9, 1–15
44 Võ, M.L.-H. and Henderson, J.M. (2010) The time course of initial scene processing for eye movement guidance in natural scene search. *J. Vis.* 10, 1–13
45 Bar, M. (2004) Visual objects in context. *Nat. Rev. Neurosci.* 5, 617–629
46 Biederman, I. *et al.* (1982) Scene perception: detecting and judging objects undergoing relational violations. *Cognit. Psychol.* 14, 143–177
47 Malcolm, G.L. and Henderson, J.M. (2009) Combining top-down processes to guide eye movements during real-world scene search. *J. Vis.* 10, 1–11
48 Hollingworth, A. (2006) Scene and position specificity in visual memory for objects. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 58–69
49 Neider, M.B. and Zelinsky, G.J. (2006) Scene context guides eye movements during visual search. *Vision Res.* 46, 614–621
50 Ehinger, K.A. *et al.* (2009) Modelling search for people in 900 scenes: a combined source model of eye guidance. *Vis. Cogn.* 17, 945–978
51 Henderson *et al.* (2009) Searching in the dark: cognitive relevance drives attention in real-world scenes. *Psychon. Bull. Rev.* 16, 850–856
52 Henderson, J.M. (2007) Regarding scenes. *Curr. Dir. Psychol. Sci.* 16, 219–222
53 Castelhano, M.S. and Henderson, J.M. (2007) Initial scene representations facilitate eye movement guidance in visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 753–763
54 Greene, M.R. and Oliva, A. (2009) Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cognit. Psychol.* 58, 137–176
55 Rousselet, G. *et al.* (2005) How long to get to the 'gist' of real-world natural scenes? *Vis. Cogn.* 12, 852–877
56 Sanocki, T. (2003) Representation and perception of scenic layout. *Cognit. Psychol.* 47, 43–86
57 Oliva, A. and Torralba, A. (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vis.* 42, 145–175

58 Biederman, I. *et al.* (1974) On the information extracted from a glance at a scene. *J. Exp. Psychol. Hum. Percept. Perform.* 103, 597–600
59 Greene, M.R. and Oliva, A. (2009) The briefest of glances: the time course of natural scene understanding. *Psychol. Sci.* 20, 464–472
60 Joubert, O.R. *et al.* (2007) Processing scene context: fast categorization and object interference. *Vision Res.* 47, 3286–3297
61 Evans, K.K. and Treisman, A. (2005) Perception of objects in natural scenes: is it really attention free? *J. Exp. Psychol. Hum. Percept. Perform.* 31, 1476–1492
62 Joubert, O.R. *et al.* (2008) Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *J. Vis.* 8, 11–111
63 Held, R. (1970) Two modes of processing spatially distributed visual stimulation. In *The Neurosciences: Second Study Program* (Schmitt, F.O., ed.), pp. 317–324, MIT Press
64 Ungerleider, L.G. and Mishkin, M. (1982) Two cortical visual systems. In *Analysis of Visual Behavior* (Ingle, D.J. *et al.*, eds), pp. 549–586, MIT Press
65 Chong, S.C. and Treisman, A. (2003) Representation of statistical properties. *Vision Res.* 43, 393–404
66 Parkes, L. *et al.* (2001) Compulsory averaging of crowded orientation signals in human vision. *Nat. Neurosci.* 4, 739–744
67 Chubb, C. *et al.* (2007) The three dimensions of human visual sensitivity to first-order contrast statistics. *Vision Res.* 47, 2237–2248
68 Williams, D.W. and Sekuler, R. (1984) Coherent global motion percepts from stochastic local motions. *Vision Res.* 24, 55–62
69 Demeyere, N. *et al.* (2008) Automatic statistical processing of visual properties in simultanagnosia. *Neuropsychologia* 46, 2861–2864
70 Alvarez, G.A. and Oliva, A. (2008) The representation of simple ensemble visual features outside the focus of attention. *Psychol. Sci.* 19, 392–398
71 Melcher, D. and Kowler, E. (1999) Shapes, surfaces and saccades. *Vision Res.* 39, 2929–2946
72 Haberman, J. and Whitney, D. (2007) Rapid extraction of mean emotion and gender from sets of faces. *Curr. Biol.* 17, R751–R753
73 Vanrullen, R. (2009) Binding hardwired versus on-demand feature conjunctions. *Vis. Cogn.* 17, 103–119
74 Krupinski, E.A. (2010) Current perspectives in medical image perception. *Atten. Percept. Psychophys.* 72, 1205–1217
75 Kundel, H.L. and Nodine, C.F. (1975) Interpreting chest radiographs without visual search. *Radiology* 116, 527–532
76 Rensink, R.A. (2000) The dynamic representation of scenes. *Vis. Cogn.* 7, 17–42
77 Chelazzi, L. *et al.* (1993) A neural basis for visual search in inferior temporal cortex. *Nature* 363, 345–347
78 Buschman, T.J. and Miller, E.K. (2009) Serial, covert shifts of attention during visual search are reflected by the frontal eye fields and correlated with population oscillations. *Neuron* 63, 386–396
79 Woodman, G.F. and Luck, S.J. (2003) Serial deployment of attention during visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 121–138
80 Bichot, N.P. *et al.* (2005) Parallel and serial neural mechanisms for visual search in macaque area V4. *Science* 308, 529–1529
81 Takeda, Y. and Yagi, A. (2000) Inhibitory tagging in visual search can be found if search stimuli remain visible. *Percept. Psychophys.* 62, 927–934
82 Klein, R. (2009) On the control of attention. *Can. J. Exp. Psychol.* 63, 240–252
83 Horowitz, T.S. and Wolfe, J.M. (1998) Visual search has no memory. *Nature* 394, 575–577
84 Klein, R.M. and MacInnes, W.J. (1999) Inhibition of return is a foraging facilitator in visual search. *Psychol. Sci.* 10, 346–1346
85 Kunar, M.A. *et al.* (2008) The role of memory and restricted context in repeated visual search. *Percept. Psychophys.* 70, 314–328
86 Wolfe, J.M. *et al.* (2000) Postattentive vision. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 693–716
87 Hollingworth, A. and Henderson, J.M. (2002) Accurate visual memory for previously attended objects in natural scenes. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 113–136
88 Jiang, Y. and Wagner, L.C. (2004) What is learned in spatial contextual cuing: configuration or individual locations? *Percept. Psychophys.* 66, 454–463
89 Neisser, U. (1967) *Cognitive psychology*, Appleton-Century-Crofts
90 Deutsch, J.A. and Deutsch, D. (1963) Attention: some theoretical considerations. *Psychol. Rev.* 70, 80–90