nature human behaviour

REVIEW ARTICLE

PUBLISHED: 8 MARCH 2017 | VOLUME: 1 | ARTICLE NUMBER: 0058

Five factors that guide attention in visual search

Jeremy M. Wolfe[1]* and Todd S. Horowitz[2]

How do we find what we are looking for? Even when the desired target is in the current field of view, we need to search because fundamental limits on visual processing make it impossible to recognize everything at once. Searching involves directing attention to objects that might be the target. This deployment of attention is not random. It is guided to the most promising items and locations by five factors discussed here: bottom-up salience, top-down feature guidance, scene structure and meaning, the previous history of search over timescales ranging from milliseconds to years, and the relative value of the targets and distractors. Modern theories of visual search need to incorporate all five factors and specify how these factors combine to shape search behaviour. An understanding of the rules of guidance can be used to improve the accuracy and efficiency of socially important search tasks, from security screening to medical image perception.

How can a texting pedestrian walk right into a pole, even though it is clearly visible[1]? At any given moment, our attention and eyes are focused on some aspects of the scene in front of us, while other portions of the visible world go relatively unattended. We deploy this selective visual attention because we are unable to fully process everything in the scene at the same time. We have the impression of seeing everything in front of our eyes, but over most of the visual field, we are probably seeing something like visual textures, rather than objects[2,3]. Identifying specific objects and apprehending their relationships to each other typically requires attention, as our unfortunate texting pedestrian can attest.

Figure 1 illustrates this point. It is obvious that this image is filled with the letters M and W in various combinations of red, blue, and yellow, but it takes attentional scrutiny to determine whether or not there is a red and yellow M.

The need to attend to objects in order to recognize them raises a problem. At any given moment, the visual field contains a very large, possibly uncountable number of objects. We can count the M and W characters of Fig. 1, but imagine looking at your reflection in the mirror. Are you an object? What about your eyes or nose or that small spot on your chin? If object recognition requires attention, and if the number of objects is uncountable, how do we manage to get our attention to a target object in a reasonable amount of time? Attention can process items at a rate of, perhaps, 20–50 items per second. If you were looking for a street sign in an urban setting containing a mere 1,000 possible objects (every window, tyre, door handle, piece of trash, and so on), it would take 20–50 seconds just to find that sign. It is introspectively obvious that you routinely find what you are looking for in the real world in a fraction of that time. To be sure, there are searches of the needle-in-a-haystack, *Where's Waldo?* variety that take significant time, but routine searches for the salt shaker, the light switch, your pen, and so forth, obviously proceed much more quickly. Search is not overwhelmed by the welter of objects in the world because search is guided to a (often very small) subset of all possible objects by several sources of information. The purpose of this article is to briefly review the growing body of knowledge about the nature of that guidance.

We will discuss five forms of guidance:

- Bottom-up, stimulus-driven guidance in which the visual properties of some aspects of the scene attract more attention than others.
- Top-down, user-driven guidance in which attention is directed to objects with known features of desired targets.
- Scene guidance in which attributes of the scene guide attention to areas likely to contain targets.
- Guidance based on the perceived value of some items or features.
- Guidance based on the history of prior search.

Measuring guidance

We can operationalize the degree of guidance in a search for a target by asking what fraction of all items can be eliminated from consideration. One of the more straightforward methods to do this is to present observers with visual search displays like those in Fig. 2 and measure the reaction time (RT) required for them to report whether or not there is a target (here a T) as a function of the number of

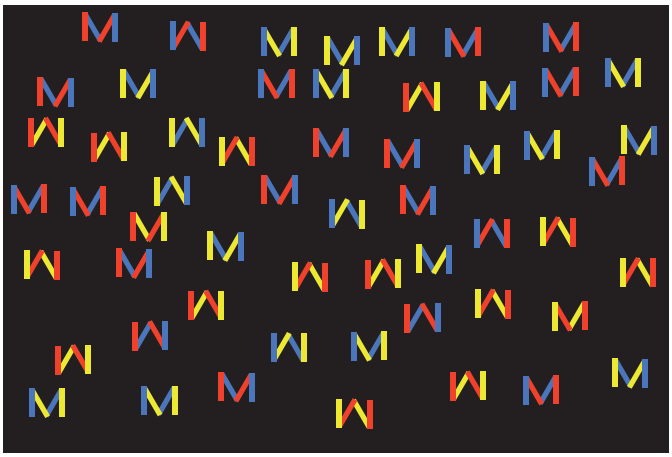

Figure 1 | A surprisingly difficult search task. On first glimpse, you know something about the distribution of colours and shapes but not how those colours and shapes are bound to each other. Find instances of the letter M that are red and yellow.

[1]Visual Attention Lab, Department of Surgery, Brigham and Women's Hospital, 64 Sidney Street, Suite 170, Cambridge, Massachusetts 02139-4170, USA. [2]Basic Biobehavioral and Psychological Sciences Branch, Behavioral Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, 9609 Medical Center Drive, 3E-116, Rockville, Maryland 20850, USA. *e-mail: jwolfe@partners.org

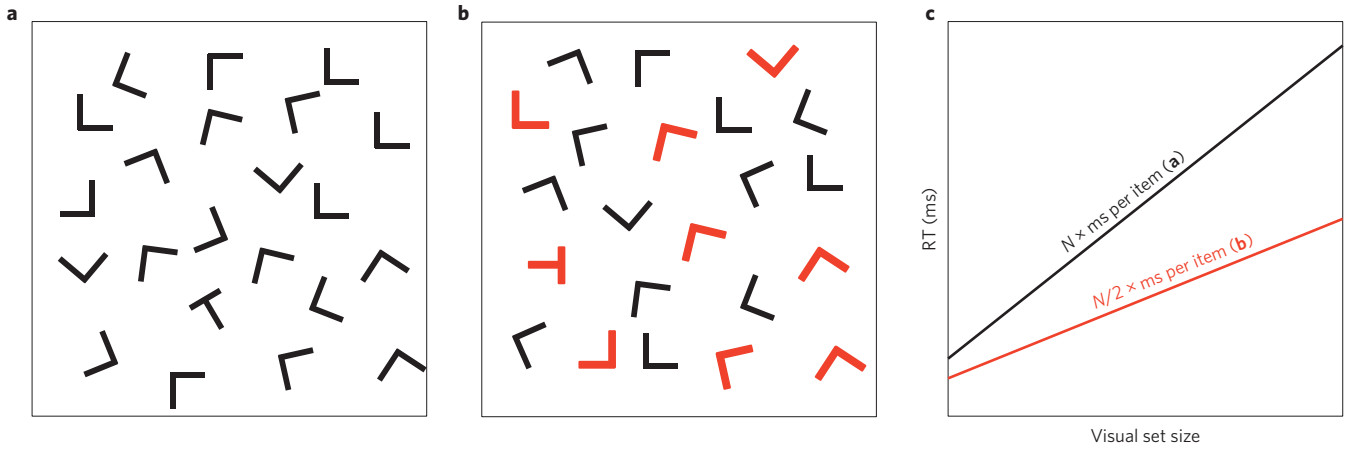

Figure 2 | The basic visual search paradigm. a–c, A target (here the letter T) is presented amidst a variable number of distractors (**a**,**b**). Search 'efficiency' can be indexed by the slope of the function relating reaction time (RT) to the visual set size (*N*) (**c**). If the target in panel **b** is a red T, the slope for **b** (red line) will be half of that for panel **a** (black line) because attention can be limited to just half of the items in **b**.

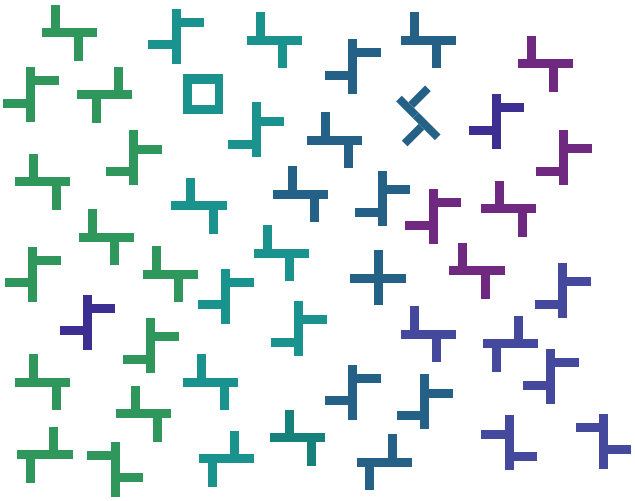

Figure 3 | Which items pop out of this display, and why?

items (set size). The slope of the RT × set size function is a measure of the efficiency of search. For a search for a T among Ls (Fig. 2a), the slope would be in the vicinity of 20–50 ms per item[4]. We believe that this reflects serial deployment of attention from item to item[5], although this need not be the case[6].

In Fig. 2b, the target is a red T. This search would be faster and more efficient[7] because attention can be guided to the red items. If half the items are red (and if guidance is perfect), the slope will be reduced by about half, suggesting that, at least in this straightforward case, slopes index the amount of guidance.

The relationship of slopes to guidance is not entirely simple, even for arrays of items like those in Fig. 2 (ref. [8]), but see ref. [9]. Matters become far more complex with real-world scenes where the visual set size is not easily defined[10,11]. However, if the slope is cut in half when half the items acquire some property, like the colour red in Fig. 2b, it is reasonable to assert that search has been guided by that property[9].

The problem of distractor rejection. As shown in Fig. 2, a stimulus attribute can make search slopes shallower by limiting the number of items in a display that need to be examined. However, guidance of attention is not the only factor that can modulate search slopes. If observers are attending to each item in the display (in series or in parallel), the slope of the RT × set size function can also be altered by changing how long it takes to reject each distractor. Thus, if we markedly reduced the contrast of Fig. 2a, the RT × set size function would become steeper, not because of a change in guidance but because it would now take longer to decide if any given item was a T or an L.

Bottom-up guidance by stimulus salience

Attention is attracted to items that differ from their surroundings, if those differences are large enough and if those differences occur in one of a limited set of attributes that guide attention. The basic principles are illustrated in Fig. 3.

Three items pop out of this display. The purple item on the left differs from its neighbours in colour. It is identical to the purple item just inside the upper right corner of the image. That second, purple item on the right is not particularly salient even though it is the only other item in that shade of purple; its neighbours are close enough in colour that the differences in colour do not attract attention. The bluish item to its immediate left is salient by virtue of an orientation difference. The square item a bit further to the left is salient because of the presence of a 'closure' feature[12] or the absence of a collection of line terminations[13]. We call properties like colour, orientation, or closure basic (or guiding) features, because they can guide the deployment of attention. Other properties may be striking when one is directly attending to an item, and may be important for object recognition, but they do not guide attention. For example, the one plus symbol in the display is not salient, even though it possesses the only X-intersection in the display, because intersection type is not a basic feature[14]. The 'pop-out' we see in Fig. 3 is not just subjective phenomenology. Pop-out refers to extremely effective guidance, and is diagnosed by a near-zero slope of the RT × set size function; although there may be systematic variability even in these 'flat' slopes[15].

There are two fundamental rules of bottom-up salience[16]. Salience of a target increases with difference from the distractors (target–distractor heterogeneity) and with the homogeneity of the distractors (distractor–distractor homogeneity) along basic feature dimensions. Bottom-up salience is the most extensively modelled aspect of visual guidance (reviewed in ref. [17]). The seminal modern work on bottom-up salience is Koch and Ullman's[18] description of a winner-take-all network for deploying attention. Subsequent decades have seen the development of several influential bottom-up models, for examples, see refs [19,20–22]. However, bottom-up salience is just one of

the factors guiding attention. By itself, it does only modestly well in predicting the deployment of attention (usually indexed by eye fixations). Models do quite well at predicting search for salience, but not as well at predicting search for other sorts of targets[17]. This is quite reasonable. If you are looking for your cat in the bedroom, it would be counterproductive to have your attention visit all the shiny, colourful objects first. Thus, a bottom-up saliency model will not do well if the observer has a clear top-down goal[23]. One might think that bottom-up salience would dominate if observers 'free-viewed' a scene in the absence of such a goal, but bottom-up models can be poor at predicting fixations even when observers free view scenes without specific instructions[24]. It seems that observers generate their own, idiosyncratic tasks, allowing other guiding forces to come into play. It is worth noting that salience models work better if they are not based purely on local features but acknowledge the structure of objects in the field of view[25]. For instance, while the most salient spot in an image might be the edge between the cat's tail and the white sheet on the bed, fixations are more likely to be directed to middle of the cat[26,27].

Top-down feature guidance

Returning to Fig. 1, if you search for Ws with yellow elements, you can guide your attention to yellow items and subsequently determine if they are Ws or Ms[7]. This is feature guidance, sometimes referred to as feature-based attention[28]. Importantly, it is possible to guide attention to more than one feature at a time. Thus, searching for a big, red, vertical item can benefit from our knowledge of its colour, size, and orientation[29]. Following the target–distractor heterogeneity rule, search efficiency is dependent on the number of features shared by targets and distractors[29], and observers appear to be able to guide attention to multiple target features simultaneously[30]. This finding raises the attractive possibility that searching for an arbitrary object among other arbitrary objects would be quite efficient because objects would be represented sparsely in a high-dimensional space. Such sparse coding has been invoked to explain object recognition[31,32]. However, searching for arbitrary objects turns out not to be particularly efficient[11,33]. By itself, guidance to multiple features does not appear to be an adequate account of how we search for objects in the real world (see 'Guidance by scene properties' section).

What are the guiding attributes?

Feature guidance bears some metaphorical similarity to your favourite computer search engine. You enter some terms into the search box and an ordered list of places to attend is returned. A major difference between internet search engines and the human visual search engine is that human search uses only a very small vocabulary of search terms (that is, features). The idea that there might be a limited set of features that could be appreciated 'preattentively'[34] was at the heart of Treisman's feature-integration theory[35]. She proposed that targets defined by unique features would pop out of displays. Subsequently, others modified the role of features to propose that they could guide the deployment of attention[7,36].

There are probably only two dozen attributes that can guide attention. The visual system can detect and identify a vast number of stimuli, but it cannot use arbitrary properties to guide attention in the way that Google or Bing can use arbitrary search terms. A list of guiding attributes is found in Box 1. This article does not list all of the citations that support each entry. Many of these can be found in older versions of the list[37,38]. Recent changes to the list are marked in italic in Box 1 and citations are given for those.

Attributes like colour are deemed to be 'undoubted' because multiple experiments from multiple labs attest to their ability to guide attention. 'Probable' feature dimensions may be merely probable because we are not sure how to define the feature. Shape is the most notable entry here. It seems quite clear that something about

Box 1 | The guiding attributes for feature search.

Changes to previous versions of the list[37,38] are marked in italics.

Undoubted guiding attributes

- Colour
- Motion
- Orientation
- Size (including length, spatial frequency, *and apparent size*[120])

Probable guiding attributes

- Luminance onset (flicker) *but see ref.* [121]
- Luminance polarity
- Vernier offset
- Stereoscopic depth and tilt
- Pictorial depth cues *but see ref.* [62]
- Shape
- Line termination
- Closure
- Curvature
- Topological status

Possible guiding attributes

- Lighting direction (shading)
- Expansion/looming
- Number
- Glossiness (lustre)
- Aspect ratio
- *Eye of origin/binocular rivalry*

Doubtful cases

- Novelty
- Letter identity alphanumeric category
- *Familiarity — over-learned sets, in general*[111]

Probably not guiding attributes

- Intersection
- Optic flow
- Colour change
- 3D volumes (for example, geons)
- Luminosity
- Material type
- Scene category
- Duration
- Stare-in-crowd[122,123]
- Biological motion
- Your name
- Threat
- Semantic category (animal, artefact, and so on)
- *Blur*[124]
- *Visual rhythm*[125]
- *Animacy/chasing*[44]
- *Threat*[45]

Faces are a complicated issue

- Faces among other objects
- Familiar faces
- Emotional faces
- Schematic faces

Factors that modulate search

- Cast shadows
- Amodal completion
- Apparent depth

shape guides attention[39]. It is less clear exactly what that might be, although the success of deep learning algorithms in enabling computers to classify objects may open up new vistas in the understanding of human search for shape[40].

The attributes described as possible await more research. Often these attributes only have a single paper supporting their entry on the list, as in the case of numerosity: can you direct attention to the pile with more elements in it, once you eliminate size, density, and other confounding visual factors? Perhaps[41], but it would be good to have converging evidence. Search for the magnitude of a digit (for example, 'find the highest number') is not guided by the semantic meaning of the digits but by their visual properties[42].

The list of attributes that do not guide attention is, of course, potentially infinite. Box 1 lists a few plausible candidates that have been tested and found wanting. For example, there has been considerable interest recently in what could be called evolutionarily motivated candidates for guidance. What would enhance our survival if we could find it efficiently? Looking at a set of moving dots on a computer screen, we can perceive that one is chasing another[43]. However, this aspect of animacy does not appear to be a guiding attribute[44]. Nor does threat (defined by association with electric shock) seem to guide search[45].

Some caution is needed here because a failure to guide attention is a negative finding and it is always possible that, were the experiment done correctly, the attribute might guide after all. Thus, early research[46] found that binocular rivalry and eye-of-origin information did not guide attention, but more recent work[47,48] suggests that it may be possible to guide attention to interocular conflict, and our own newer data[49] indicates that rivalry may guide attention if care is taken to suppress other signals that interfere with that guidance. Thus, binocular rivalry was listed under 'doubtful cases and probable non-features' in ref. [37], but is now listed under 'possible guiding attributes' in Box 1.

Faces remain a problematic candidate for feature status, with a substantial literature yielding conflicting results and conclusions. Faces are quite easy to find among other objects[50,51] but there is dispute about whether the guiding feature is 'face-ness' or some simpler stimulus attribute[52,53]. A useful review by Frischen *et al.*[54] argues that "preattentive search processes are sensitive to and influenced by facial expressions of emotion", but this is one of the cases where it is hard to reject the hypothesis that the proposed feature is modulating the processing of attended items, rather than guiding the selection of which items to attend. Suppose that, once attended, it takes 10 ms longer to disengage attention from an angry face than from a neutral face. The result would be that search would go faster (10 ms per item faster) when the distractors were neutral than when they were angry. Consequently, an angry target among neutral distractors would be found more efficiently than a neutral face among angry. Evidence for guidance by emotion would be stronger if the more efficient emotion searches were closer to pop-out than to classic inefficient, unguided searches, for example, for a T among L characters[55]. Typically, this is not the case. For example, Gerritsen *et al.*[56] report that "visual search is not blind to emotion" but, in a representative finding, search for hostile faces produced an inefficient slope of 64 ms per item even if it is somewhat more efficient than the 82 ms per item for peaceful target faces.

There are stimulus properties that, while they may not be guiding attributes in their own right, do modulate the effectiveness of other attributes. For example, apparent depth modulates apparent size, and search is guided by that apparent size[57]. Finally, there are properties of the display that influence the deployment of attention. These could be considered aspects of 'scene guidance' (see 'Guidance by scene properties' section). For example, attention tends to be attracted to the centre of gravity in a display[58]. Elements like arrows direct attention even if they themselves do not pop out[59]. As discussed by Rensink[60], these and related factors can inform graphic design and other situations where the creator of an image wants to control how the observer consumes that image.

There have been some general challenges to the enterprise of defining specific features, notably the hypothesis that many of the effects attributed to the presence or absence of basic features are actually produced by crowding in the periphery[3]. For example, is efficient search for cubes lit from one side among cubes lit from another side evidence for preattentive processing of 3D shape and lighting[61], or merely a by-product of the way these stimuli are represented in peripheral vision[62]? Resolution of this issue requires a set of visual search experiments with stimuli that are uncrowded. This probably means using low set sizes as in, for example, the evidence that material type is not a guiding attribute[63].

A different challenge to the preattentive feature enterprise is the possibility that too many discrete features are proposed. Perhaps many specific features form a continuum of guidance by a single, more broadly defined attribute. For instance, the cues to the 3D layout of the scene include stereopsis, shading, linear perspective and more. These might be part of a single attribute describing the 3D disposition of an object. Motion, onsets, and flicker might be part of a general dynamic change property[64]. Most significantly, we might combine the spatial features of line termination, closure, topological status, orientation, and so on into a single shape attribute with properties defined by the appropriate layer of the right convolutional neural net (CNN). Such nets have shown themselves capable of categorizing objects, so one could imagine a preattentive CNN guiding attention to objects as well[65]. So far, such an idea remains a promissory note. Regardless of how powerful CNNs may become, humans cannot guide attention to entirely arbitrary/specific properties in order to find particular types of object[66] and it is unknown if some intermediate representation in a CNN could capture the properties of the human search engine. If it did, we might well find that such a layer represented a space with dimensions corresponding to attributes like size, orientation, line termination, vernier offset, and so on, but this remains to be seen.

Guidance by scene properties

While the field of visual search has largely been built on search for targets in arbitrary 2D arrays of items, most real-world search takes place in structured scenes, and this structure provides a source of guidance. To illustrate, try searching for any humans in Fig. 4. Depending on the resolution of the image as you are viewing it, you may or may not be able to see legs poking out from behind the roses by the gate. Regardless, what should be clear is that the places you looked were strongly constrained. Biederman, Mezzanotte, and Rabinowitz[67] suggested a distinction between semantic and syntactic guidance.

Syntactic guidance has to do with physical constraints. You don't look for people on the front surface of the wall or in the sky because people typically need to be supported against gravity. Semantic guidance refers to the meaning of the scene. You don't look for people on the top of the wall, not because they could not be there but because they are unlikely to be there given your understanding of the scene, whereas you might scrutinize the bench. Scene guidance would be quite different (and less constrained) if the target were a bird. The use of semantic and syntactic language should not be seen as tying scene processing too closely to linguistic processing nor should the two categories be seen as neatly non-overlapping[68,69]. Nevertheless, the distinction between syntactic and semantic factors, as roughly defined here, can be observed in electrophysiological recordings: scenes showing semantic violations (for example, a bar of soap sitting next to the computer on the desk) produce different neural signatures than scenes showing syntactic violations (for example, a computer mouse on top of the laptop screen)[70]. While salience may have some influence in this task[71], it does not appear to be the major force guiding attention here[24,72]. But note that feature

guidance and scene guidance work together. People certainly could be on the lawn, but you do not scrutinize the empty lawn in Fig. 4 because it lacks the correct target features.

Extending the study of guidance from controlled arrays of distinct items to structured scenes poses some methodological challenges. For example, how do we define the set size of a scene? Is the rose bush an item in Fig. 4, or does each bloom count as an item? In bridging between the world of artificial arrays of items and scenes, perhaps the best we can do is to talk about the 'effective set size'[10,73], the number of items/locations that are treated as candidate targets in a scene given a specific task. If you are looking for the biggest flower, each rose bloom is part of the effective set. If you are looking for a human, those blooms are not part of the set. While any estimate of effective set size is imperfect, it is a very useful idea and it is clear that, for most tasks, the effective set size will be much smaller than the set of all possible items[11].

Preview methods have been very useful in examining the mechanisms of scene search[74]. A scene is flashed for a fraction of a second and then the observer searches for a target. The primary data are often eye-tracking records. It is common for these experiments to involve searching while the observer's view of the scene is restricted to a small region around the point of fixation ('gaze-contingent' displays). Very brief exposures (50–75 ms) can guide deployment of the eyes once search begins[75]. A preview of the specific scene is much more useful than a preview of another scene of the same category, although the preview scene does not need to be the same size as the search stimulus[74]. Importantly, the preview need not contain the target in order to be effective[76]. Search appears to be more strongly guided by a relatively specific scene 'gist'[73,77], an initial understanding of the scene that does not rely on recognizing specific objects[78]. The gist includes both syntactic (for example, spatial layout) and semantic information, and this combination can provide powerful search guidance. Knowledge about the target provides an independent source of guidance[79,80]. These sources of information provide useful 'priors' on where targets might be ("if there is a vase present, it's more likely to be on a table than in the sink"), which are more powerful than memory for where a target might have been seen[81,82,83] in terms of guiding search.

Preview effects may be fairly limited in search of real scenes. If the observer searches a fully visible scene rather than being limited to a gaze-contingent window, guidance by the preview is limited to the first couple of fixations[84]. Once search begins, guidance is presumably updated based on the real scene, rendering the preview obsolete. In gaze-contingent search, the effects last longer because this updating cannot occur. This updating can be seen in the work of Hwang *et al.*[85], where, in the course of normal search, the semantic content of the current fixation in a scene influences the target of the next fixation.

Modulation of search by prior history

In this section, we summarize evidence showing that the prior history of the observer, especially the prior history of search, modulates the guidance of attention. We can organize these effects by their timescale, from within a trial (on the order of 100 s to ms) to lifetime learning (on the order of years).

A number of studies have demonstrated the preview benefit: when half of the search array is presented a few hundred milliseconds before the rest of the array, the effective set size is reduced, either because attention is guided away from the old 'marked' items (visual marking[86]) and/or toward the new items (onset prioritization[87]).

On a slightly longer timescale, priming phenomena are observed from trial to trial within an experiment, and can be observed over seconds to weeks. The basic example is 'priming of pop-out'[88], in which an observer might be asked to report the shape of the one item of unique colour in a display. If that item is the one red shape among green on one trial, responses will be faster if the next trial repeats red among green as compared with a switch to green among red; although the search in both cases will be a highly efficient, colour pop-out search. More priming of pop-out is found if the task is harder[89]. Note that it is not the response, or the reporting feature, that is repeated in priming of pop-out, but the target-defining or selection feature.

Figure 4 | Scene guidance. Where is attention guided if you are looking for humans? What if the target was a bird?

More generally, seeing the features of the target makes search faster than reading a word cue describing the target, even for overlearned targets. This priming by target features takes about 200 ms to develop[90]. Priming by the features of a prior stimulus can be entirely incidental; simply repeating the target from trial to trial is sufficient[91]. More than one feature can be primed at the same time[91,92]and both target and distractor features can be primed[91,93]. Moreover, it is not just that observers are more ready to report targets with the primed feature; priming actually boosts sensitivity (that is, d')[94]. Such priming can last for at least a week[95].

Observers can also incidentally learn information over the course of an experiment that can guide search. In contextual cueing[96], a subset of the displays are repeated across several blocks of trials. While observers do not notice this repetition, RTs are faster for repeated displays than for novel, unrepeated displays[97]. The contextual cueing effect is typically interpreted as an abstract form of scene guidance: just as you learn that, in your friend's kitchen, the toaster is on the counter next to the coffee maker, you learn that, in a configuration of rotated Ls, the T is in the bottom left corner. However, evidence for this interpretation is mixed. RT × set size slopes are reduced for repeated displays[96] in some experiments, but not in others[98]. Contextual cueing effects can also be observed in cases where guidance is already nearly perfect, such as pop-out search[99] and attentionally-cued search[100]. Kunar *et al.*[98] suggested that contextual cueing reflects response facilitation, rather than guidance. Again, the evidence is mixed. There is a shift towards a more liberal response criterion for repeated displays[101], but this is not correlated with the size of the contextual cueing RT effect. In pop-out search, sensitivity to the target improves for repeated displays without an

effect on decision criterion[99]. It seems likely that observed contextual cueing effects reflect a combination of guidance effects and response facilitation, with the mix depending on the specifics of the task. Oculomotor studies show that the context is often not retrieved and available to guide attention until a search has been underway for several fixations[102,103]. Thus, the more efficient the search, the greater the likelihood that the target will be found before the context can be retrieved. Indeed, in simple letter displays, search does not become more efficient even when the same display is repeated several hundred times[104], presumably because searching *de novo* is always faster than waiting for context to become available. Once the task becomes more complex (for example, searching for that toaster)[105], it becomes worthwhile to let memory guide search[82,106].

Over years and decades, we become intimately familiar with, for example, the characters of our own written language. There is a long-running debate about whether familiarity (or, conversely, novelty) might be a basic guiding attribute. Much of this work has been conducted with overlearned categories like letters. While the topic is not settled, semantic categories like 'letter' probably do not guide attention[107,108], although mirror-reversed letters may stand out against standard letters[109,110]. Instead, items made familiar in long-term memory can modulate search[111,112], although there are limits on the effects of familiarity in search[113,114].

Modulation of search by the value of items

In the past few years, there has been increasing interest in the effects of reward or value on search. Value proves to be a strong modulator of guidance. For instance, if observers are rewarded more highly for red items than for green, they will subsequently guide attention toward red, even if this is irrelevant to the task[115]. Note, colour is the guiding feature; value modulates its effectiveness. The learned associations of value do not need to be task-relevant or salient in order to have their effects[116] and learning can be very persistent with value-driven effects being seen half a year after acquisition[117]. Indeed, the effects of value may be driving some effects of long-term familiarity described in the previous paragraph[111].

Visual search is mostly effortless. Unless we are scrutinizing aerial photographs for hints to North Korea's missile programme, or hunting for signs of cancer in a chest radiograph, we typically find what we are looking for in seconds or less. This remarkable ability is the result of attentional guidance mechanisms. While thirty-five years or so of research has given us a good grasp of the mechanisms of bottom-up salience, top-down feature-driven guidance and how those factors combine to guide attention[118,119], we are just beginning to understand how attention is guided by the structure of scenes and the sum of our past experiences. Future challenges for the field will include understanding how discrete features might fit together in a continuum of guidance and extending our theoretical frameworks from two-dimensional scenes to immersive, dynamic, three-dimensional environments.

Received 11 October 2016; accepted 27 January 2016; published 8 March 2017

References

1. Hyman, I. E., Boss, S. M., Wise, B. M., McKenzie, K. E. & Caggiano, J. M. Did you see the unicycling clown? Inattentional blindness while walking and talking on a cell phone. *Appl. Cognitive Psych.* **24,** 597–607 (2010).
2. Keshvari, S. & Rosenholtz, R. Pooling of continuous features provides a unifying account of crowding. *J. Vis.* **16,** 39 (2016).
3. Rosenholtz, R., Huang, J. & Ehinger, K. A. Rethinking the role of top-down attention in vision: effects attributable to a lossy representation in peripheral vision. *Front. Psychol.* http://dx.doi.org/10.3389/fpsyg.2012.00013 (2012).
4. Wolfe, J. M. What do 1,000,000 trials tell us about visual search? *Psychol. Sci.* **9,** 33–39 (1998).
5. Moran, R., Zehetleitner, M., Liesefeld, H., Müller, H. & Usher, M. Serial vs. parallel models of attention in visual search: accounting for benchmark RT-distributions. *Psychon. B. Rev.* **23,** 1300–1315 (2015).
6. Townsend, J. T. & Wenger, M. J. The serial-parallel dilemma: a case study in a linkage of theory and method. *Psychon. B. Rev.* **11,** 391–418 (2004).
7. Egeth, H. E., Virzi, R. A. & Garbart, H. Searching for conjunctively defined targets. *J. Exp. Psychol. Human* **10,** 32–39 (1984).
8. Kristjansson, A. Reconsidering visual search. *i-Perception* http://dx.doi.org/10.1177/2041669515614670 (2015).
9. Wolfe, J. M. Visual search revived: the slopes are not that slippery: a comment on Kristjansson (2015). *i-Perception* http://dx.doi.org/10.1177/2041669516643244 (2016).
10. Neider, M. B. & Zelinsky, G. J. Exploring set size effects in scenes: identifying the objects of search. *Vis. Cogn.* **16,** 1–10 (2008).
11. Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I. & Sherman, A. M. Visual search for arbitrary objects in real scenes. *Atten. Percept. Psychophys.* **73,** 1650–1671 (2011).
12. Kovacs, I. & Julesz, B. A closed curve is much more than an incomplete one: effect of closure in figure-ground segmentation. *Proc. Natl Acad. Sci. USA* **90,** 7495–7497 (1993).
13. Taylor, S. & Badcock, D. Processing feature density in preattentive perception. *Percept. Psychophys.* **44,** 551–562 (1988).
14. Wolfe, J. M. & DiMase, J. S. Do intersections serve as basic features in visual search? *Perception* **32,** 645–656 (2003).
15. Buetti, S., Cronin, D. A., Madison, A. M., Wang, Z. & Lleras, A. Towards a better understanding of parallel visual processing in human vision: evidence for exhaustive analysis of visual information. *J. Exp. Psychol. Gen.* **145,** 672–707 (2016).
16. Duncan, J. & Humphreys, G. W. Visual search and stimulus similarity. *Psychol. Rev.* **96,** 433–458 (1989).
17. Koehler, K., Guo, F., Zhang, S. & Eckstein, M. P. What do saliency models predict? *J. Vis.* **14,** 14 (2014).
18. Koch, C. & Ullman, S. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiol.* **4,** 219–227 (1985).
19. Itti, L., Koch, C. & Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE T. Pattern Anal.* **20,** 1254–1259 (1998).
20. Itti, L. & Koch, C. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision. Res* **40,** 1489–1506 (2000).
21. Bruce, N. D. B., Wloka, C., Frosst, N., Rahman, S. & Tsotsos, J. K. On computational modeling of visual saliency: examining what's right, and what's left. *Vision Res.* **116,** 95–112 (2015).
22. Zhang, L., Tong, M. H., Marks, T. K., Shan, H. & Cottrell, G. W. SUN: A Bayesian framework for saliency using natural statistics. *J. Vis.* **8,** 1–20 (2008).
23. Henderson, J. M., Malcolm, G. L. & Schandl, C. Searching in the dark: cognitive relevance drives attention in real-world scenes. *Psychon. Bull. Rev.* **16,** 850–856 (2009).
24. Tatler, B. W., Hayhoe, M. M., Land, M. F. & Ballard, D. H. Eye guidance in natural vision: reinterpreting salience. *J. Vis.* **11,** 5 (2011).
25. Nuthmann, A. & Henderson, J. M. Object-based attentional selection in scene viewing. *J. Vis.* **10,** 20 (2010).
26. Einhäuser, W., Spain, M. & Perona, P. Objects predict fixations better than early saliency. *J. Vis.* **8,** 18 (2008).
27. Stoll, J., Thrun, M., Nuthmann, A. & Einhäuser, W. Overt attention in natural scenes: objects dominate features. *Vision Res.* **107,** 36–48 (2015).
28. Maunsell, J. H. & Treue, S. Feature-based attention in visual cortex. *Trends Neurosci.* **29,** 317–322 (2006).
29. Nordfang, M. & Wolfe, J. M. Guided search for triple conjunctions. *Atten. Percept. Psychophys.* **76,** 1535–1559 (2014).
30. Friedman-Hill, S. R. & Wolfe, J. M. Second-order parallel processing: visual search for the odd item in a subset. *J. Exp. Psychol. Human* **21,** 531–551 (1995).
31. Olshausen, B. A. & Field, D. J. Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* **14,** 481–487 (2004).
32. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How does the brain solve visual object recognition? *Neuron* **73,** 415–434 (2012).
33. Vickery, T. J., King, L.-W. & Jiang, Y. Setting up the target template in visual search. *J. Vis.* **5,** 8 (2005).
34. Neisser, U. *Cognitive Psychology* (Appleton-Century-Crofts, 1967).
35. Treisman, A. & Gelade, G. A feature-integration theory of attention. *Cognitive Psychol.* **12,** 97–136 (1980).
36. Wolfe, J. M., Cave, K. R. & Franzel, S. L. Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Human* **15,** 419–433 (1989).
37. Wolfe, J. M. in *Oxford Handbook of Attention* (eds Nobre, A. C & Kastner, S.) 11–55 (Oxford Univ. Press, 2014).
38. Wolfe, J. M. & Horowitz, T. S. What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.* **5,** 495–501 (2004).
39. Alexander, R. G., Schmidt, J. & Zelinsky, G. J. Are summary statistics enough? Evidence for the importance of shape in guiding visual search. *Vis. Cogn.* **22,** 595–609 (2014).

40. Yamins, D. L. K. & DiCarlo, J. J. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* **19,** 356–365 (2016).
41. Reijnen, E., Wolfe, J. M. & Krummenacher, J. Coarse guidance by numerosity in visual search. *Atten. Percept. Psychophys.* **75,** 16–28 (2013).
42. Godwin, H. J., Hout, M. C. & Menneer, T. Visual similarity is stronger than semantic similarity in guiding visual search for numbers. *Psychon. Bull. Rev.* **21,** 689–695 (2014).
43. Gao, T., Newman, G. E. & Scholl, B. J. The psychophysics of chasing: a case study in the perception of animacy. *Cogn. Psychol.* **59,** 154–179 (2009).
44. Meyerhoff, H. S., Schwan, S. & Huff, M. Perceptual animacy: visual search for chasing objects among distractors. *J. Exp Psychol. Human* **40,** 702–717 (2014).
45. Notebaert, L., Crombez, G., Van Damme, S., De Houwer, J. & Theeuwes, J. Signals of threat do not capture, but prioritize, attention: a conditioning approach. *Emotion* **11,** 81–89 (2011).
46. Wolfe, J. M. & Franzel, S. L. Binocularity and visual search. *Percept. Psychophys.* **44,** 81–93 (1988).
47. Paffen, C., Hooge, I., Benjamins, J. & Hogendoorn, H. A search asymmetry for interocular conflict. *Atten. Percept. Psychophys.* **73,** 1042–1053 (2011).
48. Paffen, C. L., Hessels, R. S. & Van der Stigchel, S. Interocular conflict attracts attention. *Atten. Percept. Psychophys.* **74,** 251–256 (2012).
49. Zou, B., Utochkin, I. S., Liu, Y. & Wolfe, J. M. Binocularity and visual search—revisited. *Atten. Percept. Psychophys.* **79,** 473–483 (2016).
50. Hershler, O. & Hochstein, S. At first sight: a high-level pop out effect for faces. *Vision Res.* **45,** 1707–1724 (2005).
51. Golan, T., Bentin, S., DeGutis, J. M., Robertson, L. C. & Harel, A. Association and dissociation between detection and discrimination of objects of expertise: evidence from visual search. *Atten. Percept. Psychophys.* **76,** 391–406 (2014).
52. VanRullen, R. On second glance: still no high-level pop-out effect for faces. *Vision Res.* **46,** 3017–3027 (2006).
53. Hershler, O. & Hochstein, S. With a careful look: still no low-level confound to face pop-out. *Vision Res.* **46,** 3028–3035 (2006).
54. Frischen, A., Eastwood, J. D. & Smilek, D. Visual search for faces with emotional expressions. *Psychol. Bull.* **134,** 662–676 (2008).
55. Dugué, L., McLelland, D., Lajous, M. & VanRullen, R. Attention searches nonuniformly in space and in time. *Proc. Natl Acad. Sci. USA* **112,** 15214–15219 (2015).
56. Gerritsen, C., Frischen, A., Blake, A., Smilek, D. & Eastwood, J. D. Visual search is not blind to emotion. *Percept. Psychophys.* **70,** 1047–1059 (2008).
57. Aks, D. J. & Enns, J. T. Visual search for size is influenced by a background texture gradient. *J. Exp. Psychol. Human* **22,** 1467–1481 (1996).
58. Richards, W. & Kaufman, L. 'Centre-of-gravity' tendencies for fixations and flow patterns. *Percept. Psychophys* **5,** 81–84 (1969).
59. Kuhn, G. & Kingstone, A. Look away! Eyes and arrows engage oculomotor responses automatically. *Atten. Percept. Psychophys.* **71,** 314–327 (2009).
60. Rensink, R. A. in *Human Attention in Digital Environments* (ed. Roda, C.) Ch 3, 63–92 (Cambridge Univ. Press, 2011).
61. Enns, J. T. & Rensink, R. A. Influence of scene-based properties on visual search. *Science* **247,** 721–723 (1990).
62. Zhang, X., Huang, J., Yigit-Elliott, S. & Rosenholtz, R. Cube search, revisited. *J. Vis.* **15,** 9 (2015).
63. Wolfe, J. M. & Myers, L. Fur in the midst of the waters: visual search for material type is inefficient. *J. Vis.* **10,** 8 (2010).
64. Kunar, M. A. & Watson, D. G. Visual search in a multi-element asynchronous dynamic (MAD) world. *J. Exp. Psychol. Human* **37,** 1017–1031 (2011).
65. Ehinger, K. A. & Wolfe, J. M. How is visual search guided by shape? Using features from deep learning to understand preattentive "shape space". In *Vision Sciences Society 16th Annual Meeting* (2016); http://go.nature.com/2l1azoy
66. Vickery, T. J., King, L. W. & Jiang, Y. Setting up the target template in visual search. *J. Vis.* **5,** 81–92 (2005).
67. Biederman, I., Mezzanotte, R. J. & Rabinowitz, J. C. Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychol.* **14,** 143–177 (1982).
68. Henderson, J. M. Object identification in context: the visual processing of natural scenes. *Can. J. Psychol.* **46,** 319–341 (1992).
69. Henderson, J. M. & Hollingworth, A. High-level scene perception. *Annu. Rev. Psychol.* **50,** 243–271 (1999).
70. Vo, M. L. & Wolfe, J. M. Differential ERP signatures elicited by semantic and syntactic processing in scenes. *Psychol. Sci.* **24,** 1816–1823 (2013).
71. 't Hart, B. M., Schmidt, H. C. E. F., Klein-Harmeyer, I. & Einhäuser, W. Attention in natural scenes: contrast affects rapid visual processing and fixations alike. *Philos. T. Roy. Soc. B* **368,** http://dx.doi.org/10.1098/rstb.2013.0067 (2013).
72. Henderson, J. M., Brockmole, J. R., Castelhano, M. S. & Mack, M. L. in *Eye Movement Research: Insights into Mind and Brain* (eds van Gompel, R., Fischer, M., Murray, W., & Hill, R.) 537–562 (Elsevier, 2007).
73. Rensink, R. A. Seeing, sensing, and scrutinizing. *Vision Res.* **40,** 1469–1487 (2000).
74. Castelhano, M. S. & Henderson, J. M. Initial scene representations facilitate eye movement guidance in visual search. *J. Exp. Psychol. Human* **33,** 753–763 (2007).
75. Vo, M. L.-H. & Henderson, J. M. The time course of initial scene processing for eye movement guidance in natural scene search. *J. Vis.* **10,** 14 (2010).
76. Hollingworth, A. Two forms of scene memory guide visual search: memory for scene context and memory for the binding of target object to scene location. *Vis. Cogn.* **17,** 273–291 (2009).
77. Oliva, A. in *Neurobiology of Attention* (eds Itti, L., Rees, G., & Tsotsos, J.) 251–257 (Academic Press, 2005).
78. Greene, M. R. & Oliva, A. The briefest of glances: the time course of natural scene understanding. *Psychol. Sci.* **20,** 464–472 (2009).
79. Castelhano, M. & Heaven, C. Scene context influences without scene gist: eye movements guided by spatial associations in visual search. *Psychon. B. Rev.* **18,** 890–896 (2011).
80. Malcolm, G. L. & Henderson, J. M. Combining top-down processes to guide eye movements during real-world scene search. *J. Vis.* **10,** 1–11 (2010).
81. Torralba, A., Oliva, A., Castelhano, M. S. & Henderson, J. M. Contextual guidance of eye movements and attention in real-world scenes: the role of global features on object search. *Psychol. Rev.* **113,** 766–786 (2006).
82. Vo, M. L. & Wolfe, J. M. When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *J. Exp. Psychol. Human* **38,** 23–41 (2012).
83. Vo, M. L.-H. & Wolfe, J. M. The role of memory for visual search in scenes. *Ann. NY Acad. Sci.* **1339,** 72–81 (2015).
84. Hillstrom, A. P., Scholey, H., Liversedge, S. P. & Benson, V. The effect of the first glimpse at a scene on eye movements during search. *Psychon. B. Rev.* **19,** 204–210 (2012).
85. Hwang, A. D., Wang, H.-C. & Pomplun, M. Semantic guidance of eye movements in real-world scenes. *Vision Res.* **51,** 1192–1205 (2011).
86. Watson, D. G. & Humphreys, G. W. Visual marking: prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychol. Rev.* **104,** 90–122 (1997).
87. Donk, M. & Theeuwes, J. Prioritizing selection of new elements: bottom-up versus top-down control. *Percept. Psychophys.* **65,** 1231–1242 (2003).
88. Maljkovic, V. & Nakayama, K. Priming of popout: I. Role of features. *Mem. Cognition* **22,** 657–672 (1994).
89. Lamy, D., Zivony, A. & Yashar, A. The role of search difficulty in intertrial feature priming. *Vision Res.* **51,** 2099–2109 (2011).
90. Wolfe, J., Horowitz, T., Kenner, N. M., Hyle, M. & Vasan, N. How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Res.* **44,** 1411–1426 (2004).
91. Wolfe, J. M., Butcher, S. J., Lee, C. & Hyle, M. Changing your mind: on the contributions of top-down and bottom-up guidance in visual search for feature singletons. *J. Exp. Psychol. Human* **29,** 483–502 (2003).
92. Kristjansson, A. Simultaneous priming along multiple feature dimensions in a visual search task. *Vision Res.* **46,** 2554–2570 (2006).
93. Kristjansson, A. & Driver, J. Priming in visual search: separating the effects of target repetition, distractor repetition and role-reversal. *Vision Res.* **48,** 1217–1232 (2008).
94. Sigurdardottir, H. M., Kristjansson, A. & Driver, J. Repetition streaks increase perceptual sensitivity in visual search of brief displays. *Vis. Cogn.* **16,** 643–658 (2008).
95. Kruijne, W. & Meeter, M. Long-term priming of visual search prevails against the passage of time and counteracting instructions. *J. Exp. Psychol. Learn.* **42,** 1293–1303 (2016).
96. Chun, M. & Jiang, Y. Contextual cuing: implicit learning and memory of visual context guides spatial attention. *Cogn. Psychol.* **36,** 28–71 (1998).
97. Chun, M. M. & Jiang, Y. Top-down attentional guidance based on implicit learning of visual covariation. *Psychol. Sci.* **10,** 360–365 (1999).
98. Kunar, M. A., Flusberg, S. J., Horowitz, T. S. & Wolfe, J. M. Does contextual cueing guide the deployment of attention? *J. Exp. Psychol. Human* **33,** 816–828 (2007).
99. Geyer, T., Zehetleitner, M. & Muller, H. J. Contextual cueing of pop-out visual search: when context guides the deployment of attention. *J. Vis.* **10,** 20 (2010).
100. Schankin, A. & Schubo, A. Contextual cueing effects despite spatially cued target locations. *Psychophysiology* **47,** 717–727 (2010).
101. Schankin, A., Hagemann, D. & Schubo, A. Is contextual cueing more than the guidance of visual-spatial attention? *Biol. Psychol.* **87,** 58–65 (2011).
102. Peterson, M. S. & Kramer, A. F. Attentional guidance of the eyes by contextual information and abrupt onsets. *Percept. Psychophys.* **63,** 1239–1249 (2001).
103. Tseng, Y. C. & Li, C. S. Oculomotor correlates of context-guided learning in visual search. *Percept. Psychophys.* **66,** 1363–1378 (2004).

104. Wolfe, J. M., Klempen, N. & Dahlen, K. Post-attentive vision. *J. Exp. Psychol. Human* **26,** 693–716 (2000).
105. Brockmole, J. R. & Henderson, J. M. Using real-world scenes as contextual cues for search. *Vis. Cogn.* **13,** 99–108 (2006).
106. Hollingworth, A. & Henderson, J. M. Accurate visual memory for previously attended objects in natural scenes. *J. Exp. Psychol. Human* **28,** 113–136 (2002).
107. Flowers, J. H. & Lohr, D. J. How does familiarity affect visual search for letter strings? *Percept. Psychophys.* **37,** 557–567 (1985).
108. Krueger, L. E. The category effect in visual search depends on physical rather than conceptual differences. *Percept. Psychophys.* **35,** 558–564 (1984).
109. Frith, U. A curious effect with reversed letters explained by a theory of schema. *Percept. Psychophys.* **16,** 113–116 (1974).
110. Wang, Q., Cavanagh, P. & Green, M. Familiarity and pop-out in visual search. *Percept. Psychophy.* **56,** 495–500 (1994).
111. Qin, X. A., Koutstaal, W. & Engel, S. The hard-won benefits of familiarity on visual search — familiarity training on brand logos has little effect on search speed and efficiency. *Atten. Percept. Psychophys.* **76,** 914–930 (2014).
112. Fan, J. E. & Turk-Browne, N. B. Incidental biasing of attention from visual long-term memory. *J. Exp. Psychol. Learn.* **42,** 970–977 (2015).
113. Huang, L. Familiarity does not aid access to features. *Psychon. B. Rev.* **18,** 278–286 (2011).
114. Wolfe, J. M., Boettcher, S. E. P., Josephs, E. L., Cunningham, C. A. & Drew, T. You look familiar, but I don't care: lure rejection in hybrid visual and memory search is not based on familiarity. *J. Exp. Psychol. Human* **41,** 1576–1587 (2015).
115. Anderson, B. A., Laurent, P. A. & Yantis, S. Value-driven attentional capture. *Proc. Natl Acad. Sci. USA* **108,** 10367–10371 (2011).
116. MacLean, M. & Giesbrecht, B. Irrelevant reward and selection histories have different influences on task-relevant attentional selection. *Atten. Percept. Psychophys.* **77,** 1515–1528 (2015).
117. Anderson, B. A. & Yantis, S. Persistence of value-driven attentional capture. *J. Exp. Psychol. Human* **39,** 6–9 (2013).
118. Moran, R., Zehetleitner, M. H., Mueller, H. J. & Usher, M. Competitive guided search: meeting the challenge of benchmark RT distributions. *J. Vis.* **13,** 24 (2013).
119. Wolfe, J. M. in *Integrated Models of Cognitive Systems* (ed. Gray, W.) 99–119 (Oxford Univ. Press, 2007).
120. Proulx, M. J. & Green, M. Does apparent size capture attention in visual search? Evidence from the Müller–Lyer illusion. *J. Vis.* **11,** 21 (2011).
121. Kunar, M. A. & Watson, D. G. When are abrupt onsets found efficiently in complex visual search? Evidence from multielement asynchronous dynamic search. *J. Exp. Psychol. Human* **40,** 232–252 (2014).
122. Shirama, A. Stare in the crowd: frontal face guides overt attention independently of its gaze direction. *Perception* **41,** 447–459 (2012).
123. von Grunau, M. & Anston, C. The detection of gaze direction: a stare-in-the-crowd effect. *Perception* **24,** 1297–1313 (1995).
124. Enns, J. T. & MacDonald, S. C. The role of clarity and blur in guiding visual attention in photographs. *J. Exp. Psychol. Human* **39,** 568–578 (2013).
125. Li, H., Bao, Y., Poppel, E. & Su, Y. H. A unique visual rhythm does not pop out. *Cogn. Process.* **15,** 93–97 (2014).

Additional information

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence should be addressed to J.M.W.

How to cite this article: Wolfe, J. M. & Horowitz, T. S. Five factors that guide attention in visual search. *Nat. Hum. Behav.* **1,** 0058 (2017).

Competing interests

J.M.W occasionally serves as an expert witness or consultant (paid or unpaid) on the applications of visual search to topics from legal disputes (for example, how could that truck have hit that clearly visible motorcycle?) to consumer behaviour (for example, how could we redesign this shelf to attract more attention to our product?).