Check for updates

Incorporating the properties of peripheral vision into theories of visual search

Alejandro Lleras, *Simona Buetti* and *Zoe Jing Xu*

Abstract | People often look for objects in their immediate environment, a behaviour known as visual search. Most of the visual signals used during search come from peripheral vision, outside the direct focus of the eyes. In this Review, we present evidence that peripheral vision is both more capable and more complex than commonly believed. We then use three benchmark findings from the visual search literature to illustrate how considering peripheral vision can improve understanding of the basic mechanisms of search. Next, we discuss theories of visual search on the basis of their treatment of peripheral processing constraints and present findings in support of theories that integrate the characteristics of peripheral vision. These findings describe the span over which peripheral vision can extract useful information, the type of information peripheral vision encodes, and how peripheral vision identifies locations that are likely to contain a search target. We end by discussing considerations for future theoretical development and recommendations for future empirical research.

Many activities in daily life require locating a specific object in the world, such as searching for a spoon to add sugar to coffee or searching for a soccer ball in a park. A central question for research is how people successfully search for specific visual targets throughout the day, a task known as visual search. Understanding visual search also provides insights into how humans perform in high-stakes searches beyond the everyday, such as when looking for tumours on medical images[1], weapons on security X-ray machines[2] or military vehicles on satellite imagery.

Visual search provides a window into fundamental cognitive operations. For instance, perceiving and locating a specific target object such as a blue soccer ball in a park (FIG. 1) requires multiple processes. One must perceive the object as a set of features (circle, pentagons, blue, white) co-occurring at the same location, and represent this object (the ball) as an entity distinct from the background (the grass). In addition, visual search often requires one to search for an object without its immediate presence. Doing so relies on a target template, or internal representation of the target object. When searching for the blue soccer ball, the template could be a mental image of the ball or a list of the ball's features (colour, shape and size). This internal representation biases how spatial attention is deployed and how one moves one's eyes during search. Spatial attention refers to the mechanism that allows one to preferentially process the visual information at one location over others. While searching for the soccer ball, the target template would guide one's eyes and attention towards regions in the park that contain features similar to the target, such as the yellow ball with a similar size and shape, the blue chairs with a similar colour or the blue ball itself.

In the laboratory, visual search is most often studied under simplified conditions to isolate a subset of the mechanisms involved. For instance, backgrounds are often uniform so that all objects are equally easy to segregate from the background, the visual display is small enough to fit on a computer monitor, and the target and non-target (distractor) objects are simplified and defined by a small set of features (for example, one colour and one shape).

In this Review, we first describe the properties of peripheral vision. We then review three foundational visual search phenomena that highlight how an understanding of peripheral vision can lead to a better understanding of visual search. We then categorize influential theories of visual search on the basis of how they incorporate peripheral vision into the mechanisms that underlie visual search. Next, we present empirical work that uses various processing characteristics of peripheral vision to accurately predict search performance. We end with considerations for future theory development and empirical study in the field.

Fundamentals of visual search

At the most basic level, visual search is constrained by where the eyes are focused (the fixation point) during initial visual analysis. Visual analysis depends on where

Department of Psychology, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

✉e-mail: alejandrolleras@gmail.com; buetti@illinois.edu

https://doi.org/10.1038/s44159-022-00097-1

Fig. 1 | **Real life visual search.** Imagine this image as the view you experience when looking for a blue soccer ball in a park. First, an internal representation (target template) of the target object is activated. This object is associated with a set of known features, such as circular, blue, white and pentagon. Visual search depends on peripheral processing to help the observer find the target among non-target (distractor) objects. The region of fixation is indicated by the end of the red line, with the rest of the image processed by peripheral vision.

information is projected onto the retina because the distribution of photoreceptors on the retina varies as a function of eccentricity, or distance from the fixation point. The distribution of photoreceptors determines the fidelity with which visual information striking the retina is initially encoded by the visual system. Distance and extent of objects in the visual field are measured using degrees of visual angle. For reference, when viewed at arm's length, one's thumbnail subtends about 1.5 degrees of visual angle.

The fovea is the central part of the retina, which processes information from a region 1.7 degrees wide around the fixation point[3] (end of red line in FIG. 1). Fixating an object and thereby pointing the fovea directly at it allows access to fine visual details because the fovea is entirely populated by cones[4], retinal photoreceptor cells that support colour processing[5–10] and high-resolution vision[11–16]. The fovea also has higher visual acuity than other areas of the retina because more neurons in the brain's visual cortex are devoted to processing foveal than peripheral information — a property referred to as cortical magnification[17–19].

The visual periphery is the area that surrounds the fovea and extends approximately 214 degrees horizontally, 70 degrees above and 80 degrees below the fovea[20–22]. The visual periphery contains a much smaller concentration of cones and is mostly populated by a different type of photoreceptor — rods[4,23]. Successful everyday visual search is quite remarkable considering that the fovea covers only about 0.01% of the visual field, with the visual periphery covering the remaining 99.99%.

Both foveal and peripheral visual processing contribute to visual search. The fovea can be used to scan the visual world to evaluate individual objects until the target is found. This serial processing often involves sequential eye movements directed to specific objects (or regions of the world) at a time. It can also occur covertly when spatial attention focuses on individual objects without moving the eyes. When inspecting objects serially, reaction time is linearly related to the number of objects present in the display. The corresponding search slope characterizes the processing rate of items during search with respect to the set size of items present, measured in milliseconds per item. On average, if more objects are present, it takes observers a set time longer per object to locate the target.

When the target is sufficiently different from the distractor objects in the display, visual search becomes easier. In these easier searches, the visual system can rely on peripheral vision to simultaneously evaluate many objects. In this way, peripheral vision directly informs the deployment of spatial attention and eye movements towards the target. This form of simultaneous evaluation of information is referred to as parallel processing. When inspecting multiple objects in parallel, reaction time is logarithmically related to the number of objects present in the display (BOX 1). This means that as the number of objects increases, the increases in reaction time become smaller and smaller.

In sum, visual search is a dynamic process that involves an interplay between parallel processing of large swaths of the visual environment and more detailed inspection of individual regions of interest.

Visual processing in the periphery

Given that peripheral vision processes the vast majority of visual information, an understanding of peripheral vision is required to understand visual search. Properties of the retina and processing in early visual brain regions determine the capabilities and limitations of peripheral vision.

The low concentration of cones in the visual periphery (relative to the fovea) has led to common misconceptions in psychology that colour vision and visual acuity are severely impaired in peripheral vision. Although there is certainly some loss of information during initial encoding by peripheral vision, this loss does not entirely compromise vision. For example, colour vision is possible in the periphery, despite low cone density. Although colour sensitivity is not as high as in the fovea (FIG. 2a), colour perception remains robust even at eccentricities as large as 50 degrees with sufficiently large stimuli (8 degrees across)[6,24]. Colour vision at up to 20 degrees of eccentricity is as good as in the fovea (for stimuli 5 degrees across)[5]. Furthermore, even though it is degraded relative to foveal acuity, visual acuity in the periphery is sufficient for object and scene recognition (FIG. 2b). Observers can recognize the type of scene (for instance, a garden or a beach scene) and recognize peripheral objects even from low-resolution images, especially when colour is present[25–30] (for a review see[10]). Similarly, when controlling for cortical magnification, motion perception and the ability to distinguish between movement of different speeds are comparable in the periphery and in the fovea[31–33].

The limitation of peripheral vision comes from how information is processed by the visual system beyond initial encoding by the retina (for a review see[34]), which is not yet fully understood. However, some general properties seem clear. For instance, the information processing rate appears to slow down with eccentricity.

Visual field
The extent that can be seen with the eyes at a given fixation point, including fovea and periphery.

Fovea
The area of the retina that processes directly fixated information, with a width defined between 1.7 and 5 degrees of visual angle.

Consequently, objects that are farther in the periphery take longer to find[35–39]. More importantly, information in the periphery appears to be combined or pooled over larger regions as a function of eccentricity, with the size increasing in a linear fashion[40,41]. Pooling regions represent visual characteristics in the most efficient and compact manner with the fewest possible neurons[42–44]. These regions overlap with each other and tile the entire visual field, defining the areas over which image features are computed.

A crucial consequence of pooling-region-mediated processing is that when multiple objects fall within the same pooling region, their properties will not be coded independently. The functional consequence is an inability to separately represent objects with all their features. Consequently, observers will not be able to clearly discriminate and recognize multiple nearby (cluttered) objects in the periphery, a phenomenon referred to as visual crowding[12,41,45,46] (for a review see[47]). If a single object is presented at a peripheral location, observers can perceive it (FIG. 2c, top row). It is only the presence of clutter that results in a failure of object recognition at that same location[10] (FIG. 2c, middle row).

Visual peripheral crowding is a common occurrence in everyday vision[48]. However, there are many conditions that can minimize or break crowding (for reviews see[21,49,50]). When objects are spaced sufficiently far apart from one another, crowding dissipates[40] (FIG. 2c, bottom row), likely because the objects are processed by different pooling regions. Furthermore, the magnitude of crowding is determined by the degree of similarity between the target and nearby objects. Crowding is reduced when the target and distractor objects differ in colour[51–54], shape and size[54,55], orientation[56–58], motion[59] or darkness relative to the background[54,60]. Because objects have multiple features, crowding will be maximal when nearby objects are similar in all their features and mitigated in situations in which objects differ in one or more features[61].

In sum, peripheral vision can often provide useful information about target location. Although crowding represents the most serious limitation in peripheral visual analysis, there are many situations in which peripheral vision can still extract sufficiently useful information to guide vision and attention towards the target[62–65]. By contrast, in some extreme situations the environment is so cluttered that peripheral analysis is unable to provide any useful guidance (as in 'Where's Waldo?' images). In such cases, observers are forced to engage in a serial visual search using foveal vision.

Benchmarks of visual search

Three benchmark findings illustrate how a modern characterization of peripheral processing improves understanding of the basic mechanisms of visual search. We discuss findings from three types of search that vary in complexity: search for a target defined by a unique visual feature (feature search); search for a target defined by two features (conjunction search); and search for a real-world object embedded in a real-world scene (scene guidance effect).

Feature search. Feature search refers to search conditions in which the target differs from the distractors along a categorical feature value[66] (such as colour, FIG. 3a). Thus, search can occur in parallel, relying on simultaneous analysis of the entire visual display by peripheral vision. Feature searches are characterized by response times that are modestly (though reliably) affected by the number of items in the display and generally result in highly accurate performance. They can also be completed by relying solely on peripheral vision, without executing any eye movements[67–69]. In fact, allowing observers to move their eyes can actually slow down feature search performance[69–71].

Even within the simplicity of feature search, the conditions that break peripheral visual crowding determine the conditions in which parallel search can occur. For instance, finding a red letter among green letters will happen in parallel across the visual field, irrespective of whether the items are cluttered and crowd each other along other feature dimensions, such as shape (FIG. 3b, left). This phenomenon occurs because the red–green colour difference breaks through the crowding created by the letters. If the same display contained only black letters (FIG. 3b, middle), crowding persists and finding a specific letter is difficult. Similarly, if the colour difference between the letters is small, such as a red target and red–orange distractors (FIG. 3b, right), crowding would also occur for colour and search would not be able to unfold in parallel. Instead, eye movements and serial processing would be required[61,72,73] (FIG. 3c, red line).

The search slope in feature search was initially portrayed as flat, with search times independent of the number of items in the display. However, later work demonstrated that feature search times increase logarithmically with set size[37,74–76] (BOX 1). Thus, during parallel

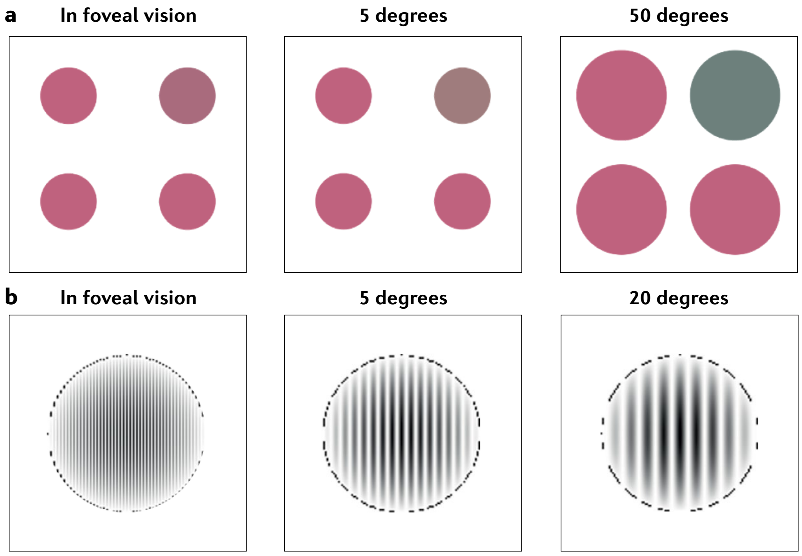
Box 1 | Logarithmic reaction times in feature search

Parallel peripheral search is characterized by a logarithmic relationship between set size and reaction time[74]. Target–distractor similarity affects the slope of the search function (the more similar target and distractors are, the steeper the function), as well as a number of other factors, including crowding, eccentricity, cortical magnification and distractor heterogeneity[37,75,76,157–159].

Parallel peripheral search is associated with stochastic, unlimited-capacity parallel processing, which determines the logarithmic shape of search slopes. Unlimited-capacity parallel processing means that peripheral vision will simultaneously process information at all locations where items are present, irrespective of the number of locations. When searching for a target, if all items take the same amount of time to be processed, processing ought to end at the same time for all items. However, visual processing is inherently stochastic: random processing fluctuations cause processing of some items to finish earlier than others. Thus, there is an additional cost to overall reaction time (RT) that comes from adding an additional item to the display.

Imagine it takes $RT_{(n)}$ to process a display with n distractors. If one more item is added to the display, $RT_{(n+1)}$ will be larger than $RT_{(n)}$ if and only if the additional item ($n+1$) is the last item to finish processing. Because processing is stochastic, all items are equally likely to be the last item to finish processing, and the probability that item ($n+1$) will be the last one to finish is $1/(n+1)$. The mathematical function that increases at the rate of $1/n$ is the natural logarithm, hence this function describes the search function for n distractors with stochastic processing.

The logarithmic relationship between reaction time and set size is found only when distractors are sufficiently different from the target that peripheral vision can confidently discriminate them from the target (as in feature search) and when observers have a fixed target template. A logarithmic relationship is not observed when the target switches from trial to trial such as in oddball search tasks[74].



c + V
+ XVK
+ X V K

Fig. 2 | **The effect of eccentricity on visual processing. a** | The minimal colour difference that can be perceived in foveal viewing (left), at 5 degrees in the periphery (centre) and at 50 degrees in the periphery (right)[6]. Note that the circles are not scaled for eccentricity. **b** | The highest spatial frequency grating, at maximum contrast, that can be barely perceived in foveal viewing (left), 5 degrees in the periphery (centre) and 20 degrees in the periphery (right). **c** | While fixating the cross in the top row, one can perceive the letter V to the right. By contrast, when fixating the cross in the middle row, the letter V is crowded by adjacent letters. When fixating the cross in the bottom row, crowding is minimized owing to the spacing of the surrounding letters.

search, adding distractors to the search display results in a measurable cost to search time. This cost occurs even when the target is quite different from the distractors in features such as colour. In addition, this cost is determined by the similarity between the target and distractors. For instance, when searching for a red target, the logarithmic function is steeper when the distractors are orange than when they are blue (FIG. 3c, orange and blue lines).

In sum, a modern characterization of peripheral processing demonstrates that feature search is more complex than initially thought. It does not rely on a categorical difference between the target and distractor features. Rather, it is a graded phenomenon sensitive to the magnitude of the target–distractor featural difference, which also determines the degree to which peripheral vision is capable of detecting the target among crowded distractors.

Conjunction search. In conjunction search, the target is defined by two feature values and shares one of these two feature values with every distractor stimulus in the display. For example, a target can be red and rectangular, whereas distractors are red triangles (same colour, different shape) or green rectangles (different colour, same shape) (FIG. 3d). Performance in conjunction search tends to be slower and more error prone than feature search (with some exceptions[77–82]) and tends to require eye movements. Thus, response times tend to significantly and linearly increase as a function of set size (FIG. 3c, red line), which is often considered an indication of serial processing (but see[83]).

It was originally thought that the higher difficulty in conjunction search was caused by an inability to represent objects defined by multiple features in peripheral vision[66]. This argument was based on two findings. First, the receptive fields (area of the visual field that a visual neuron responds to) in object-recognition brain regions such as the inferior temporal gyrus are quite large (on the order of 40 degrees of visual angle). As a consequence, precise feature location information is lost at this level of processing. Second, different visual features are processed by specialized neural feature detectors and therefore do not necessarily co-exist in a common representational space. Thus, it was proposed that the only way to find the target in conjunction search would be to inspect objects individually, directing spatial attention to an object's location to bind the features at that location into a single representation of the object[66,84,85]. Because spatial attention is a capacity-limited mechanism, it cannot be deployed at all object locations simultaneously, necessitating serial processing.

However, two processing characteristics of peripheral vision point to flaws in the traditional account of conjunction search. First, peripheral vision can bind feature information into coherent object representations under uncrowded conditions[86,87]. Thus, the size of the receptive fields is not necessarily a limitation to combining feature information. Second, pooling-mediated processing in peripheral vision can explain the inability to match features to their objects. One is not required to invoke the capacity limitations of spatial attention to explain this difficulty. Under crowded conditions, target and distractor features are processed within pooling regions that encompass multiple objects at a time[62]. Because location information is partially lost within the same pooling region, it becomes difficult to perceive whether the two features that define the target come from the same object (the target) or from two different distractor objects. Thus, the speed of conjunction search is directly determined by the ability to differentiate pooling regions that contain the target from those that do not.

Search in scenes. A third benchmark finding in visual search is the scene guidance advantage. Scenes are complex real-world visual environments that consist of surfaces, objects and backgrounds organized in a specific manner. For instance, a view of a park, a beach, a city skyline, or a kitchen can be considered a visual scene. Scenes have meaning[88,89] and structure[90] that constrain where objects tend to be located[91,92]. For instance, coffee mugs are typically found on horizontal surfaces such as kitchen counters, whereas paintings appear on vertical surfaces such as walls[89]. The scene guidance advantage is the phenomenon that when observers search for objects in scenes, attention[90] and eye movements[89] are guided towards locations that are likely to contain the target, while ignoring those that are unlikely to contain it[92].

Observers can categorize the type of scene they are looking at (such as forest, desert, or lake) extremely quickly[93–98], at a rate of up to 10 scenes per second[97,99].

They can determine basic scene properties (such as navigability, naturalness, and openness) even more quickly than that[99]. General summary statistics about objects in the scene such as their average size, orientation or expression (for faces) can also be extracted quickly and in parallel by peripheral vision, within 50 ms[100–104]. Recent theories of visual search have proposed a so-called non-selective[105] or global[89,106] pathway of visual information processing[105] to help direct attention in scenes towards likely target locations. This pathway relies on various types of fast peripheral processing that extract meaning, structure and other summary statistics from the scene to inform search[89,91,105–109].

Overall, peripheral vision can actively query a visual display for information regarding where target-like objects might be located. The success of this query is limited by visual crowding, which can be alleviated by changing where the eyes are fixated. But the quantity of information picked up by peripheral vision is nonetheless remarkable. With peripheral vision, multiple objects can be processed simultaneously and scene meaning and basic structure can be extracted to constrain the set of likely target locations.

Theoretical accounts of visual search

Theories of visual search aim to describe how observers use known information about a specific target to find it among multiple distractors in a visual display. Here, we review the most influential of these theories, organizing them into three categories that partially parallel the historical evolution of the field. For each category, we first describe how much the theories incorporate the distinction in encoding between foveal and peripheral vision. Whereas some theories are indifferent to this distinction (known as lossless theories), others include peripheral processing constraints (altered-encoding theories) or

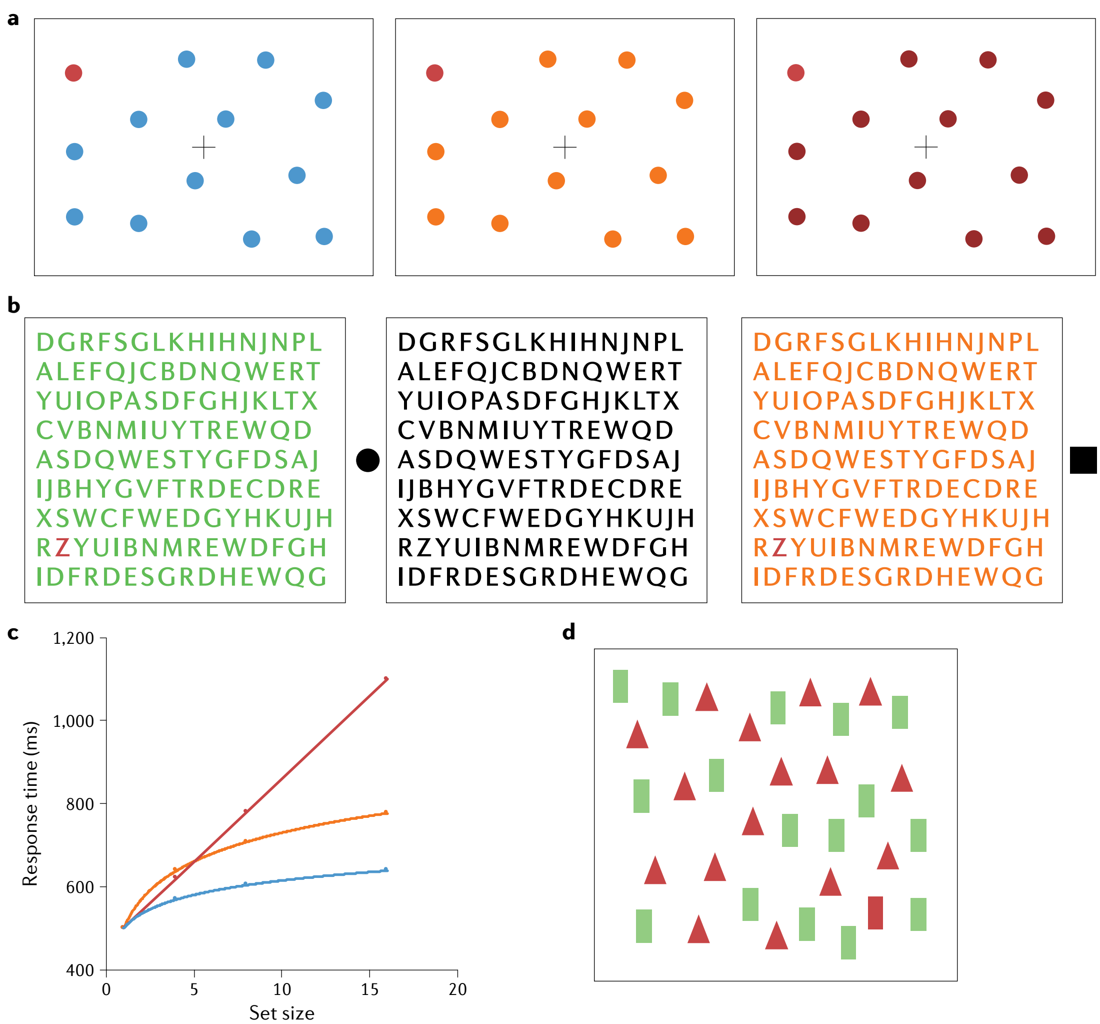


Fig. 3 | **Search conditions and search slopes. a** | Search displays with a red target among either blue, orange or dark-red distractors. **b** | While fixating the dot and therefore using peripheral vision only, it is easy to find the red Z among green letters (colour breaks the crowding) but impossible to find the black Z among black letters. Likewise, while fixating the square, it is hard to find the red Z among orange letters because of the high target–distractor similarity. **c** | Logarithmic search slopes (blue and orange lines) are observed when search is easy and unfolds in parallel, as when asked to find the red circle among blue or orange circles in panel **a**. Linear search slopes (red line) are observed when search unfolds serially, as when finding the red circle among dark-red circles in panel **a**. **d** | Conjunction search display with red rectangle target. Distractors share one feature with the target: red triangles and green rectangles.

more direct forms of periphery-limited processing. There are multiple ways in which theories can incorporate peripheral processing, such as by blurring objects farther in the periphery (to mimic low acuity), reducing the number of neurons encoding farther objects, or introducing qualitative differences between foveal and peripheral representations. Second, we discuss how various theories incorporate the search time costs associated with parallel processing.

We focus on a few key theories from each category, although many more exist (TABLE 1). Notably, theories also differ in their assumptions regarding how bottom-up (relative impact of local contrast on guidance) and top-down factors (such as similarity-based guidance, or feature boosting, whereby the representation of target features in the image are amplified to attract attention), influence search behaviour (BOX 2), which we do not discuss here.

Lossless encoding theories. Early lossless encoding theories were developed to determine what guides the eyes or spatial attention, and in what order objects will be prioritized for careful visual examination. These theories were inspired by mind-as-a-computer metaphors of information processing, and generally assume that the visual system conducts a parallel analysis of the visual display, perfectly encoding all visual features present at all eccentricities without degradation in acuity or processing. According to these accounts, visual feature dimensions are each processed in specialized brain areas (for instance, a form area and a colour area), each producing an output that registers the presence of a specific feature value (such as horizontal or red). The visual input is discussed in abstract terms, with no consideration for differences in encoding between foveal and peripheral vision[66,85,89,110–137].

Theories in this class also do not discuss the time taken for parallel processing[85,89,110,118–122,124,125,128,131–133,135,137] or they assume that parallel processing times are identical under all circumstances[66,112,126,127,129,130,134]. Instead, they characterize parallel processing as a sort of passive encoding of features that always takes the same amount of time (but see[117,123,136]). Although some theories posit that some input degradation might occur as a function of eccentricity (such as an eccentricity-dependent loss of acuity), these considerations are not directly incorporated into the search mechanisms[66,134,137].

These theories focus on the processes that occur after initial parallel analysis of the scene is completed. For instance, after analysing the features at all locations, locations are rank ordered in terms of their likelihood to contain the target. Visual search then proceeds: attention inspects each location in priority order until the target is found. Each location's priority score is independent of eye fixation, therefore the analysis and search of the scene is linked to positions within the display itself, not retinocentric positions in the visual field.

Altered-encoding theories. A second generation of theories, which we refer to as altered-encoding theories, acknowledge some differential encoding between foveal and peripheral vision. According to these theories, stimuli are blurred when processed peripherally[38,77,138–147]. This blurring only affects the information in high spatial frequency channels (fine details in the image). Other aspects, such as colour and low spatial frequency information, are not affected. Importantly, peripheral blurring increases as a function of eccentricity (eccentricity-dependent acuity loss), and is imposed before the initial analysis of the scene, before basic visual features are extracted from the image[138,147].

In some theories, peripheral-dependent encoding is also incorporated through cortical magnification, such that more eccentric locations are represented by fewer neurons (and therefore have a lower attentional weight) compared with less eccentric locations[138]. Eccentricity-dependent encoding is key to theories such as the target acquisition model[147] and the model of attention in the superior colliculus[138], which aim to predict eye movements during visual search. By contrast with the static priority maps of lossless theories[124,125,134], these theories propose that attentional maps are dynamic, with feature representations that change with every eye fixation.

A further class of theories inspired by signal detection theory re-envisioned visual search as a signal-in-noise problem[38,139,140,142–146]. The core idea is that neuronal processing, including the encoding of display features, is inherently noisy. Thus, visual search requires one to locate a noisy target signal among noisy distractor signals. Whereas lossless encoding theories code target and distractor features abstractly with unique values, signal detection theories represent target and distractor signals with a range of possible values that are normally distributed and characterized by a mean and a standard deviation. In general, the signal detection theory approach tends to study visual search at a fixed eccentricity, to control for eccentricity-mediated encoding differences. However, this framework opened the door to the idea of a representational distinction between peripheral and foveal vision. Some theories include eccentricity-dependent information loss, sometimes vaguely defined[38]. The ideal searcher model[141] operationalized eccentricity-related informational loss as a visibility map, in which the detectability of a visual feature degrades as a function of eccentricity. This degradation occurs because of both eccentricity-dependent neuronal noise and acuity loss.

In general, most altered-encoding theories focus on predicting performance measures such as accuracy or where the eyes will fixate in the scene, rather than search times. As a result, some of these theories do not discuss the time taken for parallel processing[138,141,147] or assume that it is constant[77]. That said, signal detection theories acknowledge the possibility of an eccentricity-dependent processing cost[38,139,140,142–146].

Overall, in altered-encoding theories the location of objects in the visual periphery has functional consequences in terms of how information is represented internally, which affects the likelihood of attention being directed to a location. Visual analysis in these theories is retinocentric and dynamic, changing as the eyes move. However, altered-encoding theories do not incorporate peripheral pooling regions and ignore the representational limitations associated with crowding.

Table 1 | **Visual search theories**

Category	Theory	Peripheral versus foveal encoding	Temporal cost associated with parallel processing	Bottom-up factors	Top-down factors	Key contributions
Lossless encoding theories	Two-stage model of visual search[122]	No differential encoding	None	None	Similarity to target template	Initial parallel analysis produces similarity rankings that guide serial inspection
	Feature integration theory[66]	No differential encoding	Constant	None	None	Spatial attention binds features into objects; a unique feature can attract attention
	Visual routines[137]	No differential encoding	None	Contrast-based guidance	None	Objects are represented before spatial attention; spatial attention is attracted towards salient objects
	Sagi & Julesz[129,130]	No differential encoding	Constant	Contrast-based guidance	None	Detection of feature discontinuities happens in parallel but object recognition does not
	Theory of visual attention[117]	No differential encoding	Attentional	None	Feature boosting	Similarity-based categorization and prioritization of objects
	Theeuwes[132,133]	No differential encoding	None	Contrast-based guidance	None	Unique objects capture attention
	Contingent involuntary orienting hypothesis[118]	No differential encoding	None	None	Feature boosting	Only unique objects that match the target features capture attention
	Search via recursive rejection[123]	No differential encoding	Set-size dependent	None	Similarity to target and distractor templates	Simultaneous matching to both target and distractor templates; emphasis on rejection of grouped distractors
	Guided search 2.0 (REF. 134)	No differential encoding	Constant	Contrast-based guidance	Feature boosting	Top-down signal boosts representations of one or two specific features; spatial attention inspects locations in decreasing order of priority
	Model of stimulus-driven attentional capture[121,135]	No differential encoding	None	Contrast-based guidance	Similarity to target template	New objects and objects similar to the target are prioritized to guide serial search
	Saliency-based search[124,125]	No differential encoding	None	Contrast-based guidance	None	‘Saliency map’ represents only local feature contrasts to predict movements of spatial attention
	Coherence theory of attention[85]	No differential encoding	None	None	None	Spatial attention binds features into stable representation of objects
	Optimal feature gain modulation theory[128]	No differential encoding	None	Contrast-based guidance	Optimal feature boosting	Attended feature maximizes the difference in activation between target and distractors
	Contextual guidance of eye movements[89]	No differential encoding	None	Contrast-based guidance	Spatial context guidance	Guidance towards locations where targets frequently occur
	Relational tuning[112]	No differential encoding	Constant	Contrast-based guidance	Feature relationships	Feature relations (not values) guide attention
	Signal suppression hypothesis[119,120,131]	No differential encoding	None	Contrast-based guidance	Feature down-weighting	Salient irrelevant distractors are suppressed to avoid attentional capture
	Andersen & Müller[110]	No differential encoding	None	None	Feature boosting and down-weighting	Visual selective attention first enhances target features then suppresses irrelevant distractor features
	Dimensional weighting account[126,127]	No differential encoding	Constant	Contrast-based guidance	Feature dimensions are boosted or down-weighted	Feature dimensions (not specific feature values) can be up-weighted or down-weighted
	Template shifting and asymmetrical sharpening[136]	No differential encoding	Set-size dependent	None	Feature boosting	Target template can be shifted and sharpened to increase target–distractor discriminability

Table 1 (cont.) | **Visual search theories**

Category	Theory	Peripheral versus foveal encoding	Temporal cost associated with parallel processing	Bottom-up factors	Top-down factors	Key contributions
Altered-encoding theories	Attentional engagement theory[77]	Eccentricity-dependent encoding loss	Constant	None	Similarity to target template	The combination of target–distractor and distractor–distractor similarity determines search efficiency
	Signal detection theory[38,139,140,142–146]	Eccentricity-dependent encoding loss	Eccentricity dependent	None	Target–distractor signal discriminability	Target and distractor objects have noisy representations
	Target acquisition model[147]	Eccentricity-dependent encoding loss	None	None	Similarity to target template	Predicts sequence of search fixations; similarity to the target template is dynamically re-computed after each fixation
	Ideal searcher[141]	Eccentricity-dependent encoding loss	None	None	Target–distractor signal discriminability	Fixations are planned to optimally sample information from the scene
	Model of attention in the superior colliculus[138]	Eccentricity-dependent encoding loss; cortical magnification	None	Contrast-based guidance	Feature boosting	Priority maps are projected into superior colliculus maps to predict fixations
Periphery-constrained theories	Texture tiling model[62]	Eccentricity-dependent encoding loss; differential encoding	Constant	Pooling-mediated representations	Similarity to target template	Representations of summary statistics determine search performance; spatial attention is not needed for binding features
	Buetti et al.[74]	Eccentricity-dependent encoding loss; differential encoding	Set-size dependent	None	Similarity to target template	Similarity to target template affects the time to reject unlikely targets in parallel; serial inspection of likely targets independently of similarity
	Hulleman & Olivers[148]	Eccentricity-dependent encoding loss; functional viewing field	Constant	None	Similarity to target template	Only objects falling within a functional viewing field can be recognized; field size inversely related to target–distractor similarity
	Target contrast signal theory[75]	Eccentricity-dependent encoding loss; cortical magnification; differential encoding	Set-size dependent	None	Dissimilarity to target template	Featural contrast between distractors and target template determines speed of parallel rejection; serial inspection of likely targets independent of similarity
	Guided search 6.0 (REF.[150])	Eccentricity-dependent encoding loss; functional viewing field	Constant	Contrast-based guidance	Feature boosting	Three different types of functional viewing field; scene properties can constrain deployments of attention

Prominent theories of visual search, categorized by their peripheral processing constraints and noting the contributions of bottom-up and top-down factors (BOX 2).

Periphery-constrained theories. A third set of theories focus on the conditions in which peripheral vision guides attention in a scene. For instance, one theory directly incorporates some limitations of peripheral vision as core assumptions[148]. This theory is based on the premise that only information inside a functional viewing field is analysed and can guide attention during each fixation. The functional viewing field (also known as the visual conspicuity, visual span, or useful field of view) is defined as the area beyond the foveal region where peripheral analysis is sufficient to discriminate targets from distractors without moving the eyes[149] (FIG. 4a). Importantly, the radius of the functional viewing field around fixation is dependent on target–distractor similarity because the discriminability of the target feature among the distractor features decreases with eccentricity. When target–distractor similarity is low (FIG. 4a, left), peripheral vision can differentiate between the two types of stimulus, even at large eccentricities. In this case, few or zero eye movements might be needed to find the target. When target–distractor similarity is high, the eccentricity at which peripheral vision can discriminate between stimuli is much smaller, fewer items are processed within each functional viewing field (FIG. 4a, middle and right) and more eye movements are required to eventually capture the region of the scene that contains the target. This theory proposes that what matters most for search is the number of items that can be processed within a single fixation, which in turn determines the number of times the eyes must move to find the target.

A second periphery-constrained model, guided search 6.0, further elaborated the notion of a functional viewing field by introducing three different types[150]. Specifically, if an observer is oriented towards an object and can identify it without moving their eyes, then the object falls within the visual resolution functional viewing field. Objects can be the target of a subsequent eye movement only if they are inside the exploratory functional viewing field. Finally, objects can be covertly attended (without moving the eyes) during the current fixation only if they are inside the attentional functional visual field. Thus, the existence of functional viewing fields is central to understanding search behaviour because they determine which (and how many) objects in the display can be identified, as well as the locations to which the eyes can move. Furthermore, attentional prioritization (which items are attended first) also occurs within the functional viewing field. Finally, a non-selective pathway is in charge of processing general scene information (gist of the scene, global organization of elements, summary statistics) through peripheral analysis, which can also impact how attention is deployed when search objects are embedded inside meaningful real-world scenes.

Both of these functional viewing field-focused models assume a constant time cost associated with peripheral processing. In the first theory, search times are simply determined by the number of eye movements (of constant duration) required to capture the target within the functional viewing field[148]. In guided search 6.0, the only meaningful processing times are those associated with the inspection of attended items inside the functional viewing field[150].

A third set of formal theoretical accounts explains how peripheral analysis of the scene unfolds in situations in which crowding has only a minimal effect on peripheral perception[74,75]. The core idea of the target contrast signal theory[75] is that the main function of peripheral vision is to reduce the set of locations that are likely to contain the target. When searching for a specific target, all regions in the scene that can be confidently ruled out as not containing the target are rejected in parallel so that spatial attention and eye movements have a smaller subset of objects to inspect. Objects that peripheral vision is unable to confidently categorize as distractors will be inspected in random order by spatial attention and eye movements[151], resulting in a linear search function. Notably, it is not strictly necessary for the parallel rejection and serial inspection of individual locations to proceed sequentially[152].

The parallel rejection of distractors across the visual field is modelled as a noisy evidence accumulation process that is sensitive to factors such as eccentricity, size, and target–distractor contrast. Target–distractor contrast is defined as the distance in feature space between the target feature and the distractor feature. For instance, when looking for a red target amongst either blue or orange distractor objects (FIG. 3a, left and middle), because the red–blue colour difference (contrast) is larger than the red–orange contrast, evidence for rejection will accumulate more quickly for blue than for orange distractors (FIG. 4b). Consequently, when searching for a red target, reaction times will be shorter and the logarithmic slope will be smaller in displays that contain blue items than in displays that contain orange items. Overall, the magnitude of the search slope is proposed to be inversely related to the target–distractor contrast (FIG. 4c). In terms of search speed (the inverse of search slope), when search conditions become easier (when target–distractor contrast is large), search speed is high, indicating that more objects per unit time are being processed. When search conditions become harder (when target–distractor contrast is small), search speed is low, indicating that fewer objects per unit time are being processed. Working in reverse, analysing search times provides an indirect measure of target–distractor contrast. In sum, according to the target contrast signal theory, distractor rejection is a process that takes time, even though it unfolds in parallel for some distractors.

Finally, the texture tiling model[62,153] attempts to incorporate the unique processing characteristics of peripheral vision, including crowding, into a theory of visual search. This theory proposes that objects that are processed within the same peripheral pooling region cannot be represented independently. Instead, the visual system must rely on the average summary statistics that together describe all of the objects in that region. Through pooling and summary statistics processing,

Box 2 | **Bottom-up and top-down factors in visual search**

We characterize theories in this Review with respect to peripheral processing constraints. However, most of these theories were conceived with different theoretical goals in mind and differ from one another along other key fundamental properties (TABLE 1). In particular, theories differ in their assumptions of how bottom-up and top-down factors impact search behaviour.

Bottom-up factors that influence how attention is allocated include stimulus characteristics and how stimuli are encoded by early visual brain areas. For example, a stimulus that is very different from its surroundings is more likely to capture attention (a key observation of contrast-based guidance[112,119–121,124–135,137,138,150]). But capture will also depend on where the stimulus falls in the visual field and how it is encoded and represented by early visual areas[62]. Some theories do not discuss the contribution of bottom-up factors to the allocation of attention[38,66,74,75,77,85,110,117,122,123,136,139–148] or minimize this contribution[118].

Top-down factors represent observers' goals or information used to direct attention towards stimuli that are likely to be targets. One family of theories proposes that top-down information alters the representations of stimuli by modulating the activation associated with certain target features (or feature dimensions). For instance, when looking for a red scarf, the internal representation of all red items in the scene would be enhanced to increase their attentional priority. The specifics of these modulations is variously feature boosting[110,117,118,134,136,138,150], down-weighting of distractor features[110,119,120,131], optimal boosting of the features that maximize the difference between target and distractor signals[128], spatial guidance towards target-frequent locations[89], or modulation of the importance of an entire feature dimension (such as colour)[126,127]. One theory[112] proposes that attention can be tuned to feature relationships (attention moves towards the reddest stimulus) rather than to specific feature values (such as red). Other theories emphasize quantifying the similarity of each stimulus to the target template[62,74,77,121–123,135,147,148], such that more target-similar stimuli are more likely to be attended, whereas others emphasize the discriminability of target and distractor signals[38,139–146]. Finally, one theory[75] proposes that target dissimilarity has an asymmetrical role in guidance: high levels of dissimilarity to the target template allow for the efficient rejection of unlikely targets but high levels of similarity are not used to prioritize attentional selection towards target-likely stimuli[151].

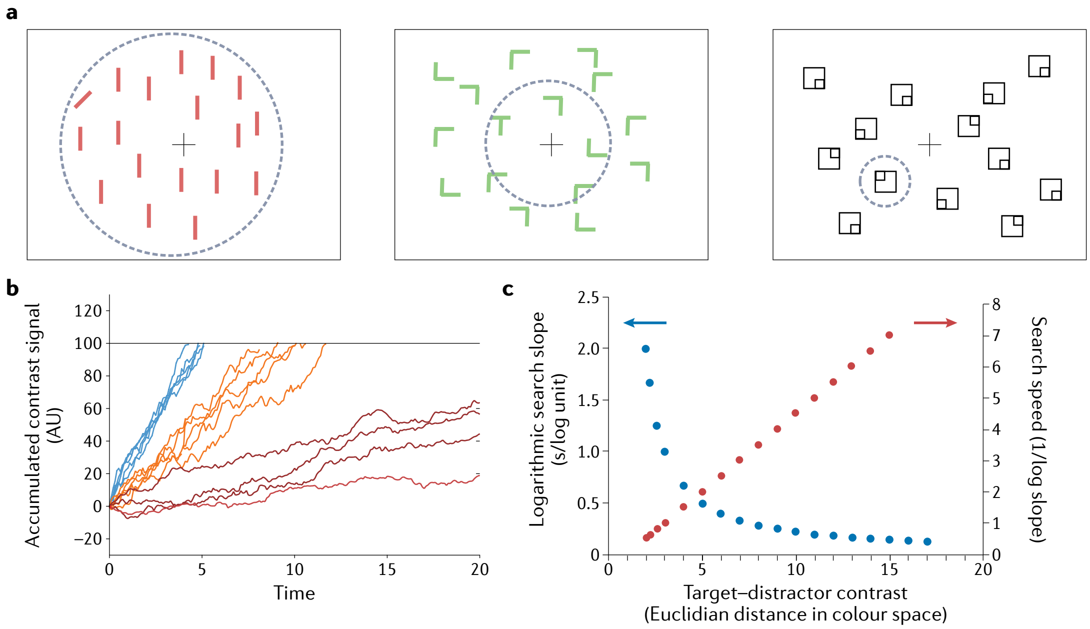


Fig. 4 | **Elements of periphery-constrained theories. a** | The functional viewing field[148] limits the range of usable information during visual search. The functional viewing field shrinks as a function of search difficulty from easiest (left) to hardest (right). **b** | Evidence accumulation for search displays in FIG. 3a, following the target contrast signal theory[75]. At every time step, evidence is increased by a constant amount (the evidence accumulation rate), plus or minus some normally distributed noise (the stochastic component). The evidence accumulation rate is proportional to the target–distractor contrast. On average, blue accumulators (lines) will reach the rejection threshold (solid horizontal line) before orange accumulators (lines) given that the red–blue distance in colour space is larger than the red–orange distance. The dark-red and red accumulators (lines) fail to reach the rejection threshold owing to lack of contrast and correspond to objects that require serial inspection. **c** | According to the target contrast signal theory, the magnitude of the logarithmic search slope is inversely proportional to the target–distractor contrast[75] (blue dots) and the search speed is linearly related to it (red dots). AU, arbitrary units.

some information about each object is lost: the specific location of each feature in the pooled area and which features belong to which object. Thus, for any given peripheral pooling region, the representation of the visual input does not consist of a list of veridical features present at the crowded location. Despite this loss of information, pooling-region-mediated summary statistics are sufficient to guide behaviour in visual search[62]. The texture tiling model is the only theory that specifies how peripheral vision changes the representation of features.

Considering the three theory classes, many of the lossless and altered-encoding theories include mechanisms that aim to explain phenomena in visual search that are out of the scope of periphery-constrained theories (such as selection history effects and attentional capture). Thus, there is room for periphery-constrained theories to broaden their explanatory power by including some of these mechanisms. In the following section, we evaluate these theories for their ability to account for empirical data in visual search.

Evaluating peripheral search theories

In this section, we review the most important findings in support of the periphery-constrained theories of visual search presented above. We focus on findings regarding functional viewing fields, summary statistics and target–distractor contrast. These findings demonstrate the progress that has been made in empirical and theoretical development and help outline where the field can move next.

Functional viewing field limits search. The proposal of a limited functional viewing field in which objects can be discriminated from each other[148] and information can be extracted to guide attention[150] is a substantial step towards integrating the characteristics of peripheral processing into working theories of attention and visual search. Evidence for the functional viewing field comes from simulations of visual search performance under three different levels of search difficulty: easy (feature search, FIG. 4a, left), medium (serial search using Ts and Ls, FIG. 4a, middle) and hard (serial search using square-within-square stimuli, FIG. 4a, right). These levels of difficulty differ in the similarity between target and distractors. Human search performance in these tasks fits well to simulated search performance from the functional viewing field theory[148], which includes a limit to the number of stimuli that can be simultaneously inspected. This limit mirrors the number of items within a hypothetical functional viewing field, the size of which is dictated by target–distractor similarity. Further evidence comes from gaze-contingent experiments in which only the items falling inside a specific area around fixation were displayed to participants[154]. In this paradigm, search performance was degraded only in the easy search conditions, suggesting that items outside of the functional field of view do not impact search behaviour in more difficult search tasks.

A similarity-dependent functional viewing field is also consistent with the texture tiling account because the likelihood that two objects crowd each other increases with their similarity[62,65]. Pooling regions increase in size with eccentricity, therefore, at a given level of target–distractor similarity, targets and distractors can be independently coded and differentiated at small eccentricities (FIG. 5a, small ellipse), yet crowd each other at farther eccentricities (FIG. 5a, large ellipse). As a consequence, there is an area around fixation where processing can unfold in parallel — the functional viewing field.

The target contrast signal theory[75] provides a framework to understand the emergence of functional viewing field-like restrictions of peripheral processing — even in simple feature search — that have been observed in human data[69]. For instance, although observers can search for a red target among blue or orange distractors without executing eye movements[69], when eye movements are allowed, observers tend to move their eyes when the distractors are orange but keep them still when they are blue. Yet, when target–distractor similarity is higher and peripheral analysis unfolds more slowly, the same observers are more likely to make an eye movement before a full examination of the most peripheral items is complete[75]. This pattern indicates a preference to execute eye movements when target–distractor similarity is increased, even when search is easy.

The functional viewing field might represent the transition point between uncrowded vision around fixation and crowded perception in the periphery, marking the boundary where pooling-mediated processing begins to replace an object-based understanding of the scene, or it might represent the area of the display where useful information was extracted from before observers move their eyes. The crucial idea is that there is a functional limit to the information that observers process on each eye fixation. Further, as proposed by guided search

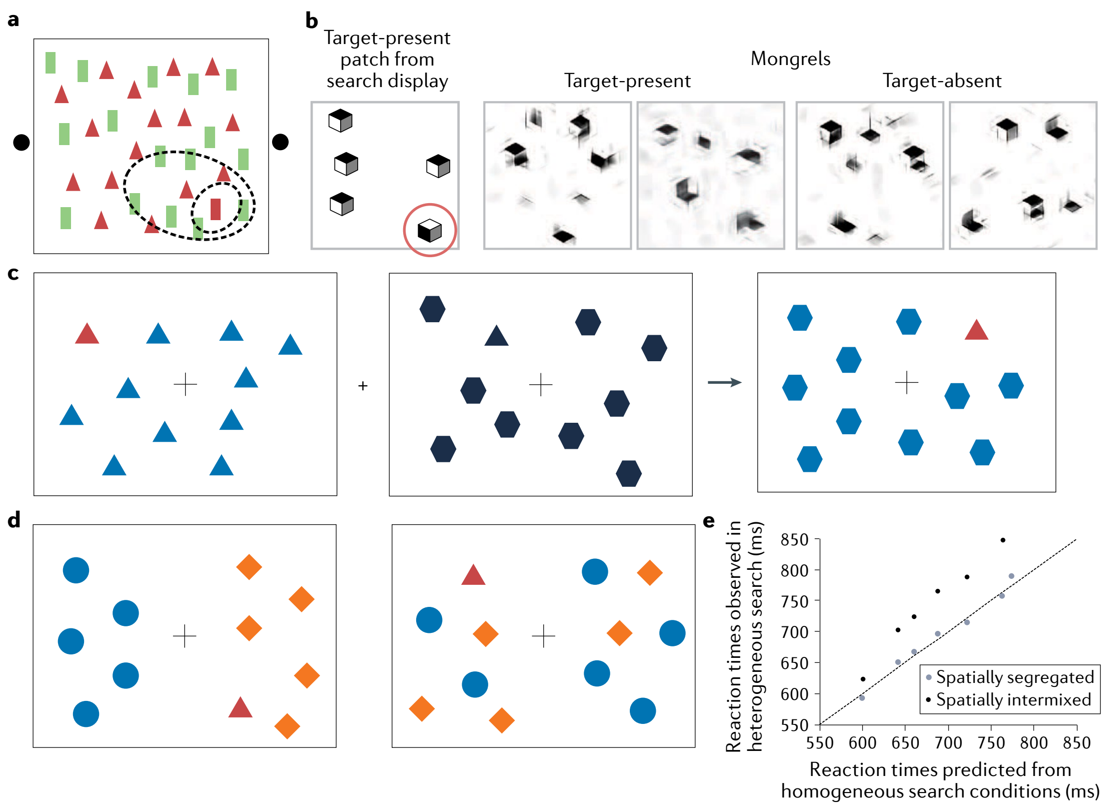


Fig. 5 | **Peripherally informed visual search. a** | Conjunction search with pooling regions. The largest ellipse visualizes a pooling region (corresponding to the left fixation dot) that contains the target. Peripheral vision will represent this region with a set of summary statistics that combines features from both types of distractor. The smaller ellipse visualizes an equivalent pooling region for the right fixation dot. At this eccentricity, pooling regions can represent information associated with a single item. The target cannot be perceived when fixating the left dot, but can be perceived when fixating the right dot. **b** | Search for a cube that is white on top (left) and corresponding mongrel images, which visualize the peripheral representation of target-present (centre) and target-absent (right) regions of the display. **c** | Target–distractor contrast for colour and shape feature combination search. Target–distractor contrast in the colour dimension (left) can be added linearly to the target–distractor contrast in the shape dimension (centre) to predict the overall target–distractor contrast when a target differs from distractors in both colour and shape (right)[86]. **d** | Local distractor–distractor similarity effects. Two types of distractor are present, either spatially segregated (left) or intermixed (right). **e** | Reaction times in the segregated display (grey circles) can be predicted using target–distractor contrasts measured in homogeneous conditions (as in panel **c**), but performance in the intermixed condition is multiplicatively slower (black circles) than in the segregated condition[157]. Panel **b** reprinted with permission from REF.[65], Journal of Vision.

6.0, different forms of processing might be limited to different extents, something that future theories should further explore.

Summary statistics predict search. A novel paradigm reveals the influence of summary statistics on search performance by comparing performance on two tasks[62,65]. First, participants conduct visual search with a range of stimuli associated with a wide spectrum of search difficulty. For instance, in a hard-difficulty task, participants are asked to find the cube with the white top (FIG. 5b, left). Second, in a separate task, participants search for a target in a display that contains the same summary statistics as the displays used in the search tasks, but with the features scrambled across the objects to create 'mongrel' images (FIG. 5b, middle and right). Participants are asked to discriminate whether the target is present or absent, given unlimited time, with the display presented at fixation. This task is considered an indirect index of peripheral discriminability, based on summary statistics of crowded stimuli. Participants' ability to discriminate between target-present and target-absent mongrel images should mimic their ability to differentiate between target-present and target-absent regions of the display when they are processed by peripheral vision during search. The search slopes from the first task are highly predictive of the performance (discriminability, d') on the second task[62,65]. This finding provides strong evidence that in difficult search — when displays are crowded and search slopes are linear — peripheral summary statistics determine search performance.

This evidence supports the use of summary statistics in the texture tiling model, which can make useful and subtle predictions regarding how small stimulus manipulations impact crowded search performance[155]. These manipulations influence the summary statistic representations that drive behaviour during search, by increasing or decreasing the confusability of regions that do or do not contain the target. Importantly, these modulations in performance cannot readily be explained by lossless encoding theories because the stimulus features themselves have not been fundamentally altered.

The goal of the texture tiling model was to demonstrate that a better understanding of crowding can lead to a better understanding of search behaviour. Converging evidence comes from a study demonstrating that the speed at which observers search a display is correlated with their individual susceptibility to crowding[156]. Susceptibility to crowding was evaluated by determining the width of critical spacing required to avoid crowding in the periphery: the larger the susceptibility to crowding, the more slowly the search unfolds.

Overall, these results stress the importance of incorporating peripheral crowding as a key determinant of search behaviour. Summary statistics such as those used in the texture tiling model could be incorporated into other models of search as a better way of understanding the representations at play under crowded conditions.

Contrast determines search speed. Target–distractor contrast is a stable measure that characterizes the time needed to reject distractors using peripheral vision (FIG. 4c) when searching for a specific target. This relationship forms the basis of support and central inspiration for the target contrast signal theory. The logarithmic search slope for a specific target–distractor pair can be used to predict search performance by novel observers and in novel experimental conditions[86,87,157–159]. We review two of these studies below, one testing the mathematical laws that underlie feature combinations[86] and the other testing the effects of intermixing different types of distractor[157].

The first relevant data regarding target–distractor contrast are from feature combination search. In a scene containing a target that differs from distractors in both colour and shape (FIG. 5c, right), it is possible to predict search times based on search performance observed under simpler conditions where the target and distractors differ only in colour (FIG. 5c, left) or shape (FIG. 5c, middle)[86]. Specifically, search speed for a target defined by a colour and a shape difference (feature combination search) is the sum of the search speed for the colour difference alone and the search speed for the shape difference alone. A similar investigation revealed that when target and distractors differ in shape and surface texture (such as dotted or striped stimuli), search performance can be predicted by a mathematical combination of the speed in simpler tasks in which target and distractors differ only in shape or texture[87]. Overall, these findings demonstrate that the target–distractor contrast is a stable measure that the visual system computes and relies on to search.

The second key piece of evidence regarding target–distractor contrast relates to distractor–distractor similarity. It has been known for decades that distractor–distractor similarity is an important factor in determining search efficiency, but it was difficult to measure[77]. However, with the realization that the logarithmic search slope is a stable index of target–distractor contrast (similarity), one can separately evaluate the distinct contribution of distractor–distractor similarity[157]. Distractor–distractor similarity affects search performance on various levels, both within and across neighbouring pooling regions.

Even for displays with the same types of distractor present overall, their spatial arrangement influences search time. When nearby distractors are identical to one another — either in a homogeneous display or when two types of distractor are presented in a spatially segregated fashion — they facilitate rejection of each other. For instance, if all identical distractors are spatially segregated on one side of the display (FIG. 5d, left), reaction times can be perfectly predicted by the slopes observed in homogeneous displays with only one kind of distractor[157] (FIG. 5e, grey circles). However, if the same distractors are spatially intermixed on the display (FIG. 5d, right), there is a systematic slow-down across all conditions relative to the homogeneous display[157] (FIG. 5e, black circles). When nearby distractors differ from one another, they are slower to reject, leading to slower reaction times[157]. These results on heterogeneous search have been replicated with stimuli of varying complexity, from simple oriented lines[159], to simple coloured geometric shapes[157], to images of real-world objects[158].

In summary, findings using a wide variety of approaches have started to characterize how visual processing unfolds in peripheral vision and how it affects search performance. Empirical data and simulations suggest that peripheral vision is able to extract useful information only over restricted regions of the display, which vary as a function of search difficulty. Further, this information might or might not be a veridical representation of objects in the world, depending on the level of crowding. Yet the output of the peripheral analysis remains useful in that it determines visual search behaviour. These findings provide strong support for periphery-constrained theories of visual search and open new avenues for further research and theoretical developments.

Summary and future directions

A contemporary understanding of peripheral processing challenges long-held assumptions about visual search. There is relatively preserved colour vision and sufficient visual acuity to recognize objects and scene properties far into the periphery. Integrating the characteristics of peripheral vision into theories of visual search has led to a better understanding of peripheral processing under crowded and uncrowded conditions, of the spatial extent over which peripheral processing can process objects in parallel, and of the temporal costs associated with peripheral scene analysis.

The ability of peripheral vision to reduce the spatial uncertainty in a scene — determining regions where the target is likely — is perhaps one of its greatest contributions to visual search behaviour in daily life. Returning to the example of searching for a soccer ball (FIG. 1), peripheral vision can accurately code and integrate the features of uncrowded objects such as the blue ball and the bench into objects that can be compared in parallel with the target template. Even the crowded pooling region containing the shrubs and trees carries enough colour and shape information to quickly determine that the blue ball is not present there. By contrast, when looking for a specific shrub using peripheral vision, crowding will obfuscate the individual features of each shrub and force one to move one's eyes to sequentially inspect small regions of the scene.

Future theories should reconsider the proposal that the visual system ranks objects in a scene in terms of their similarity to a target template and then prioritizes their processing accordingly. Under crowded conditions, peripheral vision cannot represent objects independently and therefore cannot rank them by similarity. Similarity-based ranking is further complicated by the fact that as similarity increases, crowding becomes more pronounced. Thus, spatial prioritization might not be a matter of attention being attracted towards target-similar objects or regions. Dissimilarity (or featural contrast) is probably a more useful concept. Peripheral vision can identify regions that score high in dissimilarity and reject them as unlikely to contain the target. This rejection allows for an efficient reduction in spatial uncertainty, limiting the number of locations that are likely to contain the target.

Finally, several promising lines of empirical investigation are emerging. For example, it remains to be discovered how peripheral vision segments crowded objects from their background and whether the visual information of the background is incorporated into the pooling-mediated representation. More generally, much remains unknown about visual search under visual conditions closer to the real world. Research thus far has focused on studying peripheral vision with 2D images, presented on flat computer displays. Future research should study how peripheral processing changes when the scene is three dimensional and objects appear at varying distances from the observer, and therefore require different degrees of focal accommodation by the eyes. Finally, in the real world, visual scenes are neither fully crowded nor entirely uncrowded. Local properties of a scene tend to vary by region in terms of how similar nearby objects are to one another. Future research should study how humans search in those more varied and realistic visual environments.

Published online: 01 September 2022

1. Evans, K. K., Birdwell, R. L. & Wolfe, J. M. If you don't find it often, you often don't find it: why some cancers are missed in breast cancer screening. *PLoS ONE* **8**, e64366 (2013).
2. Biggs, A. T., Cain, M. S., Clark, K., Darling, E. F. & Mitroff, S. R. Assessing visual search performance differences between Transportation Security Administration Officers and nonprofessional visual searchers. *Vis. Cognit.* **21**, 330–352 (2013).
3. Wandell, B. A. *Foundations of Vision* (Sinauer Associates, 1995).
4. Curcio, C. A., Sloan, K. R., Kalina, R. E. & Hendrickson, A. E. Human photoreceptor topography. *J. Comp. Neurol.* **292**, 497–523 (1990).
5. Abramov, I., Gordon, J. & Chan, H. Color appearance in the peripheral retina: effects of stimulus size. *J. Opt. Soc. Am. A* **8**, 404–414 (1991).
6. Hansen, T., Pracejus, L. & Gegenfurtner, K. R. Color perception in the intermediate periphery of the visual field. *J. Vis.* **9**, 26–26 (2009).
7. Martin, P. R., Lee, B. B., White, A. J., Solomon, S. G. & Rüttiger, L. Chromatic sensitivity of ganglion cells in the peripheral primate retina. *Nature* **410**, 933–936 (2001).
8. Mullen, K. T. Differential distributions of red–green and blue–yellow cone opponency across the visual field. *Vis. Neurosci.* **19**, 109–118 (2002).
9. Mullen, K. T., Sakurai, M. & Chu, W. Does L/M cone opponency disappear in human periphery? *Perception* **34**, 951–959 (2005).
10. Rosenholtz, R. Capabilities and limitations of peripheral vision. *Annu. Rev. Vis. Sci.* **2**, 437–457 (2016).
11. Anstis, S. Picturing peripheral acuity. *Perception* **27**, 817–825 (1998).
12. Anstis, S. M. A chart demonstrating variations in acuity with retinal position. *Vis. Res.* **14**, 589–592 (1974).
13. Findlay, J. M. & Gilchrist, I. D. *Active Vision: the Psychology of Looking and Seeing* (Oxford Univ. Press, 2003).
14. Levi, D. M., Klein, S. A. & Aitsebaomo, A. Vernier acuity, crowding and cortical magnification. *Vis. Res.* **25**, 963–977 (1985).
15. Rovamo, J., Leinonen, L., Laurinen, P. & Virsu, V. Temporal integration and contrast sensitivity in foveal and peripheral vision. *Perception* **13**, 665–674 (1984).
16. Wertheim, T. Über die indirekte Sehschärfe. *Z. Psychol. Physiol. Sinnesorgane* **7**, 172–187 (1894).
17. Daniel, P. & Whitteridge, D. The representation of the visual field on the cerebral cortex in monkeys. *J. Physiol.* **159**, 203–221 (1961).
18. Cowey, A. & Rolls, E. Human cortical magnification factor and its relation to visual acuity. *Exp. Brain Res.* **21**, 447–454 (1974).
19. Duncan, R. O. & Boynton, G. M. Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron* **38**, 659–671 (2003).
20. Rönne, H. Zur theorie und technik der Bjerrumschen gesichtsfelduntersuchung. *Arch. Augenheilkd.* **78**, 284–301 (1915).
21. Strasburger, H. Seven myths on crowding and peripheral vision. *i-Perception* **11**, 2041669520913052 (2020). **This paper is a comprehensive review of crowding, including biological mechanisms, common measurements and behavioural impacts.**
22. Traquair, H. M. *An Introduction to Clinical Perimetry*, Ch. 1. 4–5 (Henry Kimpton, 1938).
23. Oesterberg, G. *Topography of the Layer of Rods and Cones in the Human Retina* (NYT Nordisk Forlag: Arnold Busck, 1935).
24. Tyler, C. W. Peripheral color demo. *i-Perception* **6**, 2041669515613671 (2015).
25. Castelhano, M. S. & Henderson, J. M. The influence of color on the perception of scene gist. *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 660 (2008).
26. Oliva, A. & Schyns, P. G. Diagnostic colors mediate scene recognition. *Cognit. Psychol.* **41**, 176–210 (2000).
27. Oliva, A. & Torralba, A. Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vis.* **42**, 145–175 (2001).
28. Rousselet, G., Joubert, O. & Fabre-Thorpe, M. How long to get to the "gist" of real-world natural scenes? *Vis. Cognit.* **12**, 852–877 (2005).
29. Torralba, A. How many pixels make an image. *Vis. Neurosci.* **26**, 123–131 (2009).

30. Wurm, L. H., Legge, G. E., Isenberg, L. M. & Luebker, A. Color improves object recognition in normal and low vision. *J. Exp. Psychol. Hum. Percept. Perform.* **19**, 899 (1993).
31. McKee, S. P. & Nakayama, K. The detection of motion in the peripheral visual field. *Vis. Res.* **24**, 25–32 (1984).
32. Johnston, A. & Wright, M. Lower thresholds of motion for gratings as a function of eccentricity and contrast. *Vis. Res.* **25**, 179–185 (1985).
33. Johnston, A. & Wright, M. Matching velocity in central and peripheral vision. *Vis. Res.* **26**, 1099–1109 (1986).
34. Strasburger, H., Rentschler, I. & Jüttner, M. Peripheral vision and pattern recognition: a review. *J. Vis.* **11**, 13–13 (2011).
This paper provides a thorough review of the properties of peripheral vision.
35. Carrasco, M., Evert, D. L., Chang, I. & Katz, S. M. The eccentricity effect: target eccentricity affects performance on conjunction searches. *Percept. Psychophys.* **57**, 1241–1261 (1995).
36. Scialfa, C. T. & Joffe, K. M. Response times and eye movements in feature and conjunction search as a function of target eccentricity. *Percept. Psychophys.* **60**, 1067–1082 (1998).
37. Wang, Z., Lleras, A. & Buetti, S. Parallel, exhaustive processing underlies logarithmic search functions: visual search with cortical magnification. *Psychon. Bull. Rev.* **25**, 1343–1350 (2018).
38. Geisler, W. S. & Chou, K.-L. Separation of low-level and high-level factors in complex tasks: visual search. *Psychol. Rev.* **102**, 356 (1995).
39. Kee, D., Jung, E. S. & Chung, M. K. Isoresponse time regions for the evaluation of visual search performance in ergonomic interface models. *Ergonomics* **35**, 243–252 (1992).
40. Bouma, H. Interaction effects in parafoveal letter recognition. *Nature* **226**, 177–178 (1970).
41. Pelli, D. G., Palomares, M. & Majaj, N. J. Crowding is unlike ordinary masking: distinguishing feature integration from detection. *J. Vis.* **4**, 12–12 (2004).
42. Freeman, J. & Simoncelli, E. P. Metamers of the ventral stream. *Nat. Neurosci.* **14**, 1195–1201 (2011).
43. Gatys, L., Ecker, A. S. & Bethge, M. Texture synthesis using convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **28**, 262–270 (2015).
44. Portilla, J. & Simoncelli, E. P. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vis.* **40**, 49–70 (2000).
45. Stuart, J. A. & Burian, H. M. A study of separation difficulty: its relationship to visual acuity in normal and amblyopic eyes. *Am. J. Ophthalmol.* **53**, 471–477 (1962).
46. Ehlers, H. The movements of the eyes during reading. *Acta Ophthalmol.* **14**, 56–63 (1936).
47. Pelli, D. G. Crowding: a cortical constraint on object recognition. *Curr. Opin. Neurobiol.* **18**, 445–451 (2008).
48. Ringer, R. V., Coy, A. M., Larson, A. M. & Loschky, L. C. Investigating visual crowding of objects in complex real-world scenes. *Iperception* **12**, 2041669521994150 (2021).
49. Whitney, D. & Levi, D. M. Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends Cognit. Sci.* **15**, 160–168 (2011).
50. Herzog, M. H. & Manassi, M. Uncorking the bottleneck of crowding: a fresh look at object recognition. *Curr. Opin. Behav. Sci.* **1**, 86–93 (2015).
51. Gheri, C., Morgan, M. J. & Solomon, J. A. The relationship between search efficiency and crowding. *Perception* **36**, 1779–1787 (2007).
52. Poder, E. & Wagemans, J. Crowding with conjunctions of simple features. *J. Vis.* **7**, 23 (2007).
53. Kennedy, G. J. & Whitaker, D. The chromatic selectivity of visual crowding. *J. Vis.* **10**, 15 (2010).
54. Kooi, F. L., Toet, A., Tripathy, S. P. & Levi, D. M. The effect of similarity and duration on spatial interaction in peripheral vision. *Spat. Vis.* **8**, 255–280 (1994).
55. Nazir, T. A. Effects of lateral masking and spatial precueing on gap-resolution in central and peripheral vision. *Vis. Res.* **32**, 771–777 (1992).
56. Andriessen, J. & Bouma, H. Eccentric vision: adverse interactions between line segments. *Vis. Res.* **16**, 71–78 (1976).
57. Hariharan, S., Levi, D. M. & Klein, S. A. "Crowding" in normal and amblyopic vision assessed with Gaussian and Gabor C's. *Vis. Res.* **45**, 617–633 (2005).
58. Levi, D. M., Hariharan, S. & Klein, S. A. Suppressive and facilitatory spatial interactions in amblyopic vision. *Vis. Res.* **42**, 1379–1394 (2002).
59. Banton, T. & Levi, D. M. Spatial localization of motion-defined and luminance-defined contours. *Vis. Res.* **33**, 2225–2237 (1993).
60. Chakravarthi, R. & Cavanagh, P. Temporal properties of the polarity advantage effect in crowding. *J. Vis.* **7**, 11 (2007).
61. Greenwood, J. A. & Parsons, M. J. Dissociable effects of visual crowding on the perception of color and motion. *Proc. Natl Acad. Sci. USA* **117**, 8196–8202 (2020).
62. Rosenholtz, R., Huang, J., Raj, A., Balas, B. J. & Ilie, L. A summary statistic representation in peripheral vision explains visual search. *J. Vis.* **12**, 14 (2012).
This paper introduces a visual search theory that fully incorporates the unique processing characteristics of peripheral vision to predict crowded visual search behaviour.
63. Popple, A. V. & Levi, D. M. The perception of spatial order at a glance. *Vis. Res.* **45**, 1085–1090 (2005).
64. Strasburger, H. Unfocussed spatial attention underlies the crowding effect in indirect form vision. *J. Vis.* **5**, 8 (2005).
65. Zhang, X., Huang, J., Yigit-Elliott, S. & Rosenholtz, R. Cube search, revisited. *J. Vis.* **15**, 9 (2015).
66. Treisman, A. M. & Gelade, G. A feature-integration theory of attention. *Cognit. Psychol.* **12**, 97–136 (1980).
This paper proposes that focused attention is required for feature binding and the formation of object representations, based on feature search and conjunction search.
67. Eckstein, M. P., Beutter, B. R. & Stone, L. S. Quantifying the performance limits of human saccadic targeting during visual search. *Perception* **30**, 1389–1401 (2001).
68. Klein, R. & Farrell, M. Search performance without eye movements. *Percept. Psychophys.* **46**, 476–482 (1989).
69. Ng, G. J. P., Lleras, A. & Buetti, S. Fixed-target efficient search has logarithmic efficiency with and without eye movements. *Atten. Percept. Psychophys.* **80**, 1752–1762 (2018).
70. Carrasco, M., McLean, T. L., Katz, S. M. & Frieder, K. S. Feature asymmetries in visual search: effects of display duration, target eccentricity, orientation and spatial frequency. *Vis. Res.* **38**, 347–374 (1998).
71. Zelinsky, G. J. & Sheinberg, D. L. Eye movements during parallel–serial visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **23**, 244 (1997).
72. Carter, R. C. Visual search with color. *J. Exp. Psychol. Hum. Percept. Perform.* **8**, 127 (1982).
73. Nagy, A. L. & Sanchez, R. R. Critical color differences determined with a visual search task. *J. Opt. Soc. Am. A* **7**, 1209–1217 (1990).
74. 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 (2016).
This paper documents the search time costs associated with parallel search, which follow a logarithmic function of set size and track target–distractor similarity.
75. Lleras, A. et al. A target contrast signal theory of parallel processing in goal-directed search. *Atten. Percept. Psychophys.* **82**, 394–425 (2020).
This paper proposes that the logarithmic cost to simultaneously reject distractors is determined by target–distractor contrast in feature space.
76. Madison, A., Lleras, A. & Buetti, S. The role of crowding in parallel search: peripheral pooling is not responsible for logarithmic efficiency in parallel search. *Atten. Percept. Psychophys.* **80**, 352–373 (2018).
77. Duncan, J. & Humphreys, G. W. Visual search and stimulus similarity. *Psychol. Rev.* **96**, 433 (1989).
78. Egeth, H. E., Virzi, R. A. & Garbart, H. Searching for conjunctively defined targets. *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 32 (1984).
79. McLeod, P., Driver, J. & Crisp, J. Visual search for a conjunction of movement and form is parallel. *Nature* **332**, 154–155 (1988).
80. Nakayama, K. & Silverman, G. H. Serial and parallel processing of visual feature conjunctions. *Nature* **320**, 264–265 (1986).
81. Quinlan, P. T. & Humphreys, G. W. Visual search for targets defined by combinations of color, shape, and size: an examination of the task constraints on feature and conjunction searches. *Percept. Psychophys.* **41**, 455–472 (1987).
82. Wolfe, J. M., Cave, K. R. & Franzel, S. L. Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **15**, 419 (1989).
83. Townsend, J. T. & Ashby, F. G. *Stochastic Modeling of Elementary Psychological Processes*, 47–98 (CUP Archive, 1983).
This paper is an early systematic investigation of various possible cognitive architectures underlying visual search.
84. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
85. Rensink, R. A. The dynamic representation of scenes. *Vis. Cognit.* **7**, 17–42 (2000).
86. Buetti, S., Xu, J. & Lleras, A. Predicting how color and shape combine in the human visual system to direct attention. *Sci. Rep.* **9**, 20258 (2019).
This paper identifies the mathematical law by which colour and shape differences between targets and distractors are combined to determine the overall speed at which parallel search unfolds.
87. Xu, Z. J., Lleras, A. & Buetti, S. Predicting how surface texture and shape combine in the human visual system to direct attention. *Sci. Rep.* **11**, 6170 (2021).
88. Henderson, J. M. & Hayes, T. R. Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nat. Hum. Behav.* **1**, 743–747 (2017).
89. 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 in object search. *Psychol. Rev.* **113**, 766 (2006).
This paper describes a model that predicts eye movements during search in real-world scenes.
90. Pereira, E. J. & Castelhano, M. S. Attentional capture is contingent on scene region: using surface guidance framework to explore attentional mechanisms during search. *Psychon. Bull. Rev.* **26**, 1273–1281 (2019).
91. Neider, M. B. & Zelinsky, G. J. Scene context guides eye movements during visual search. *Vis. Res.* **46**, 614–621 (2006).
92. Võ, M. L.-H. The meaning and structure of scenes. *Vis. Res.* **181**, 10–20 (2021).
93. Greene, M. R. & Oliva, A. Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cognit. Psychol.* **58**, 137–176 (2009).
94. Loffler, G., Gordon, G. E., Wilkinson, F., Goren, D. & Wilson, H. R. Configural masking of faces: evidence for high-level interactions in face perception. *Vis. Res.* **45**, 2287–2297 (2005).
95. Loschky, L. C. et al. The importance of information localization in scene gist recognition. *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 1431 (2007).
96. Macé, M. J.-M., Joubert, O. R., Nespoulous, J.-L. & Fabre-Thorpe, M. The time-course of visual categorizations: you spot the animal faster than the bird. *PLoS ONE* **4**, e5927 (2009).
97. Potter, M. C. Meaning in visual search. *Science* **187**, 965–966 (1975).
98. Potter, M. C. & Fox, L. F. Detecting and remembering simultaneous pictures in a rapid serial visual presentation. *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 28 (2009).
99. Greene, M. R. & Oliva, A. The briefest of glances: the time course of natural scene understanding. *Psychol. Sci.* **20**, 464–472 (2009).
100. Alvarez, G. A. & Oliva, A. The representation of simple ensemble visual features outside the focus of attention. *Psychol. Sci.* **19**, 392–398 (2008).
101. Chong, S. C. & Treisman, A. Representation of statistical properties. *Vis. Res.* **43**, 393–404 (2003).
102. Chong, S. C. & Treisman, A. Attentional spread in the statistical processing of visual displays. *Percept. Psychophys.* **67**, 1–13 (2005).
103. Chong, S. C. & Treisman, A. Statistical processing: computing the average size in perceptual groups. *Vis. Res.* **45**, 891–900 (2005).
104. Haberman, J. & Whitney, D. Seeing the mean: ensemble coding for sets of faces. *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 718 (2009).
105. Wolfe, J. M., Võ, M. L.-H., Evans, K. K. & Greene, M. R. Visual search in scenes involves selective and nonselective pathways. *Trends Cognit. Sci.* **15**, 77–84 (2011).
106. Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A. & Oliva, A. Modelling search for people in 900 scenes: a combined source model of eye guidance. *Vis. Cognit.* **17**, 945–978 (2009).
107. Castelhano, M. S. & Heaven, C. The relative contribution of scene context and target features to

visual search in scenes. *Atten. Percept. Psychophys.* **72**, 1283–1297 (2010).
108. Droll, J. & Eckstein, M. Expected object position of two hundred fifty observers predicts first fixations of seventy seven separate observers during search. *J. Vis.* **8**, 320 (2008).
109. Mack, S. C. & Eckstein, M. P. Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *J. Vis.* **11**, 9 (2011).
110. Andersen, S. & Müller, M. Behavioral performance follows the time course of neural facilitation and suppression during cued shifts of feature-selective attention. *Proc. Natl Acad. Sci. USA* **107**, 13878–13882 (2010).
111. Becker, S. I. Can intertrial effects of features and dimensions be explained by a single theory? *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 1417 (2008).
112. Becker, S. I. The role of target–distractor relationships in guiding attention and the eyes in visual search. *J. Exp. Psychol. Gen.* **139**, 247 (2010).
113. Becker, S. I. Simply shapely: relative, not absolute shapes are primed in pop-out search. *Atten. Percept. Psychophys.* **75**, 845–861 (2013).
114. Becker, S. I., Folk, C. L. & Remington, R. W. Attentional capture does not depend on feature similarity, but on target-nontarget relations. *Psychol. Sci.* **24**, 634–647 (2013).
115. Becker, S. I., Harris, A. M., Venini, D. & Retell, J. D. Visual search for color and shape: when is the gaze guided by feature relationships, when by feature values? *J. Exp. Psychol. Hum. Percept. Perform.* **40**, 264 (2014).
116. Becker, S. I., Harris, A. M., York, A. & Choi, J. Conjunction search is relational: behavioral and electrophysiological evidence. *J. Exp. Psychol. Hum. Percept. Perform.* **43**, 1828 (2017).
117. Bundesen, C. A theory of visual attention. *Psychol. Rev.* **97**, 523 (1990).
118. Folk, C. L., Remington, R. W. & Johnston, J. C. Involuntary covert orienting is contingent on attentional control settings. *J. Exp. Psychol. Hum. Percept. Perform.* **18**, 1030 (1992).
119. Gaspelin, N. & Luck, S. J. Distinguishing among potential mechanisms of singleton suppression. *J. Exp. Psychol. Hum. Percept. Perform.* **44**, 626 (2018).
120. Gaspelin, N. & Luck, S. J. Inhibition as a potential resolution to the attentional capture debate. *Curr. Opin. Psychol.* **29**, 12 (2019).
121. Hillstrom, A. P. & Yantis, S. Visual motion and attentional capture. *Percept. Psychophys.* **55**, 399–411 (1994).
122. Hoffman, J. E. A two-stage model of visual search. *Percept. Psychophys.* **25**, 319–327 (1979).
123. Humphreys, G. W. & Muller, H. J. Search via recursive rejection (SERR): a connectionist model of visual search. *Cognit. Psychol.* **25**, 43–110 (1993).
124. Itti, L. & Koch, C. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vis. Res.* **40**, 1489–1506 (2000).
125. Itti, L., Koch, C. & Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 1254–1259 (1998).
126. Liesefeld, H. R., Liesefeld, A. M., Pollmann, S. & Müller, H. J. in *Processes of Visuospatial Attention and Working Memory* (ed. Hodgson, T.) 87–113 (Springer, 2018).
127. Liesefeld, H. R. & Müller, H. J. Distractor handling via dimension weighting. *Curr. Opin. Psychol.* **29**, 160–167 (2019).
128. Navalpakkam, V. & Itti, L. Search goal tunes visual features optimally. *Neuron* **53**, 605–617 (2007).
129. Sagi, D. & Julesz, B. Detection versus discrimination of visual orientation. *Perception* **13**, 619–628 (1984).
130. Sagi, D. & Julesz, B. "Where" and" what" in vision. *Science* **228**, 1217–1219 (1985).
131. Sawaki, R. & Luck, S. J. Capture versus suppression of attention by salient singletons: electrophysiological evidence for an automatic attend-to-me signal. *Atten. Percept. Psychophys.* **72**, 1455–1470 (2010).
132. Theeuwes, J. Cross-dimensional perceptual selectivity. *Percept. Psychophys.* **50**, 184–193 (1991).
133. Theeuwes, J. Perceptual selectivity for color and form. *Percept. Psychophys.* **51**, 599–606 (1992).
134. Wolfe, J. M. Guided search 2.0 a revised model of visual search. *Psychon. Bull. Rev.* **1**, 202–238 (1994).
135. Yantis, S. & Hillstrom, A. P. Stimulus-driven attentional capture: evidence from equiluminant visual objects. *J. Exp. Psychol. Hum. Percept. Perform.* **20**, 95 (1994).
136. Yu, X. & Geng, J. J. The attentional template is shifted and asymmetrically sharpened by distractor context. *J. Exp. Psychol. Hum. Percept. Perform.* **45**, 336 (2019).
137. Ullman, S. in *Readings in Computer Vision* (eds Fischler, M. A. & Firschein, O.) 298–328 (Elsevier, 1987).
138. Adeli, H., Vitu, F. & Zelinsky, G. J. A model of the superior colliculus predicts fixation locations during scene viewing and visual search. *J. Neurosci.* **37**, 1453–1467 (2017).
139. Palmer, J. & Wright, R. Attentional effects in visual search: relating search accuracy and search time. *Vis. Atten.* **8**, 348–388 (1998).
140. Eckstein, M. P., Thomas, J. P., Palmer, J. & Shimozaki, S. S. A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Percept. Psychophys.* **62**, 425–451 (2000).
141. Najemnik, J. & Geisler, W. S. Optimal eye movement strategies in visual search. *Nature* **434**, 387–391 (2005).
This paper presents a computational model of visual search that compares human performance with an ideal observer.
142. Palmer, J. Set-size effects in visual search: the effect of attention is independent of the stimulus for simple tasks. *Vis. Res.* **34**, 1703–1721 (1994).
143. Palmer, J. Attention in visual search: distinguishing four causes of a set-size effect. *Curr. Dir. Psychol. Sci.* **4**, 118–123 (1995).
144. Palmer, J., Ames, C. T. & Lindsey, D. T. Measuring the effect of attention on simple visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **19**, 108 (1993).
145. Palmer, J., Verghese, P. & Pavel, M. The psychophysics of visual search. *Vis. Res.* **40**, 1227–1268 (2000).
146. Verghese, P. Visual search and attention: a signal detection theory approach. *Neuron* **31**, 523–535 (2001).
147. Zelinsky, G. J. A theory of eye movements during target acquisition. *Psychol. Rev.* **115**, 787 (2008).
This paper describes a visual search theory focused on predicting the sequence of eye movements required to find a target in a scene.
148. Hulleman, J. & Olivers, C. N. The impending demise of the item in visual search. *Behav. Brain Sci.* **40**, e132 (2017).
This paper proposes that observers can differentiate targets from distractors only inside a functional viewing field around fixation.
149. Sanders, A. F. Some aspects of the selective process in the functional visual field. *Ergonomics* **13**, 101–117 (1970).
150. Wolfe, J. M. Guided Search 6.0: an updated model of visual search. *Psychon. Bull. Rev.* **28**, 1060–1092 (2021).
This paper presents a comprehensive model of visual search that combines the concepts of functional viewing fields and scene information with feature-based guidance.
151. Ng, G. J., Buetti, S., Patel, T. N. & Lleras, A. Prioritization in visual attention does not work the way you think it does. *J. Exp. Psychol. Hum. Percept. Perform.* **47**, 252–268 (2020).
152. Heaton, R., Hummel, J., Lleras, A. & Buetti, S. A computational account of serial and parallel processing in visual search. *J. Vis.* **20**, 844 (2020).
153. Balas, B., Nakano, L. & Rosenholtz, R. A summary-statistic representation in peripheral vision explains visual crowding. *J. Vis.* **9**, 13 (2009).
154. Young, A. H. & Hulleman, J. Eye movements reveal how task difficulty moulds visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 168 (2013).
155. Chang, H. & Rosenholtz, R. Search performance is better predicted by tileability than presence of a unique basic feature. *J. Vis.* **16**, 13 (2016).
156. Veríssimo, I. S., Hölsken, S. & Olivers, C. N. Individual differences in crowding predict visual search performance. *J. Vis.* **21**, 29–29 (2021).
157. Lleras, A. et al. Predicting search performance in heterogeneous scenes: quantifying the impact of homogeneity effects in efficient search. *Collabra Psychol.* **5**, 151 (2019).
158. Wang, Z., Buetti, S. & Lleras, A. Predicting search performance in heterogeneous visual search scenes with real-world objects. *Collabra Psychol.* **3**, 53 (2017).
159. Xu, Z. J., Lleras, A., Shao, Y. & Buetti, S. Distractor–distractor interactions in visual search for oriented targets explain the increased difficulty observed in nonlinearly separable conditions. *J. Exp. Psychol. Hum. Percept. Perform.* **47**, 1274 (2021).

Acknowledgements
The authors thank R. Rosenholtz for her feedback on the first draft of the manuscript, as well as the National Science Foundation for partially supporting this project under grant no. BCS1921735 to S.B.

Author contributions
All authors researched data for the article and contributed substantially to discussion of the content. A.L. and S.B. wrote the article. All authors reviewed and/or edited the manuscript before submission.

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Psychology thanks Jeremy Wolfe and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.