Annual Review of Vision Science

Visual Search: How Do We Find What We Are Looking For?

Jeremy M. Wolfe[1,2,3]

[1]Department of Ophthalmology, Harvard Medical School, Boston, Massachusetts 02115, USA; email: jwolfe@bwh.harvard.edu

[2]Department of Radiology, Harvard Medical School, Boston, Massachusetts 02115, USA

[3]Visual Attention Lab, Brigham & Women's Hospital, Cambridge, Massachusetts 02139, USA

Sofia Krasovskaya

VML Lab Meeting seminar

Feb 5th, 2021

This review concentrates on four topics: how visual search is guided, how serial and parallel processes collaborate in visual search, search templates and working memory (WM), and search termination and tasks beyond laboratory search.

5 factors that guide attention (Wolfe & Horowitz 2017).

1) Bottom-up (stimulus-driven saliency)

2) Top-down (user-driven): Where is the ‘template’?

3) Scene guidance

4) Value (reward-driven attention)

5) History (priming & repetition effects)

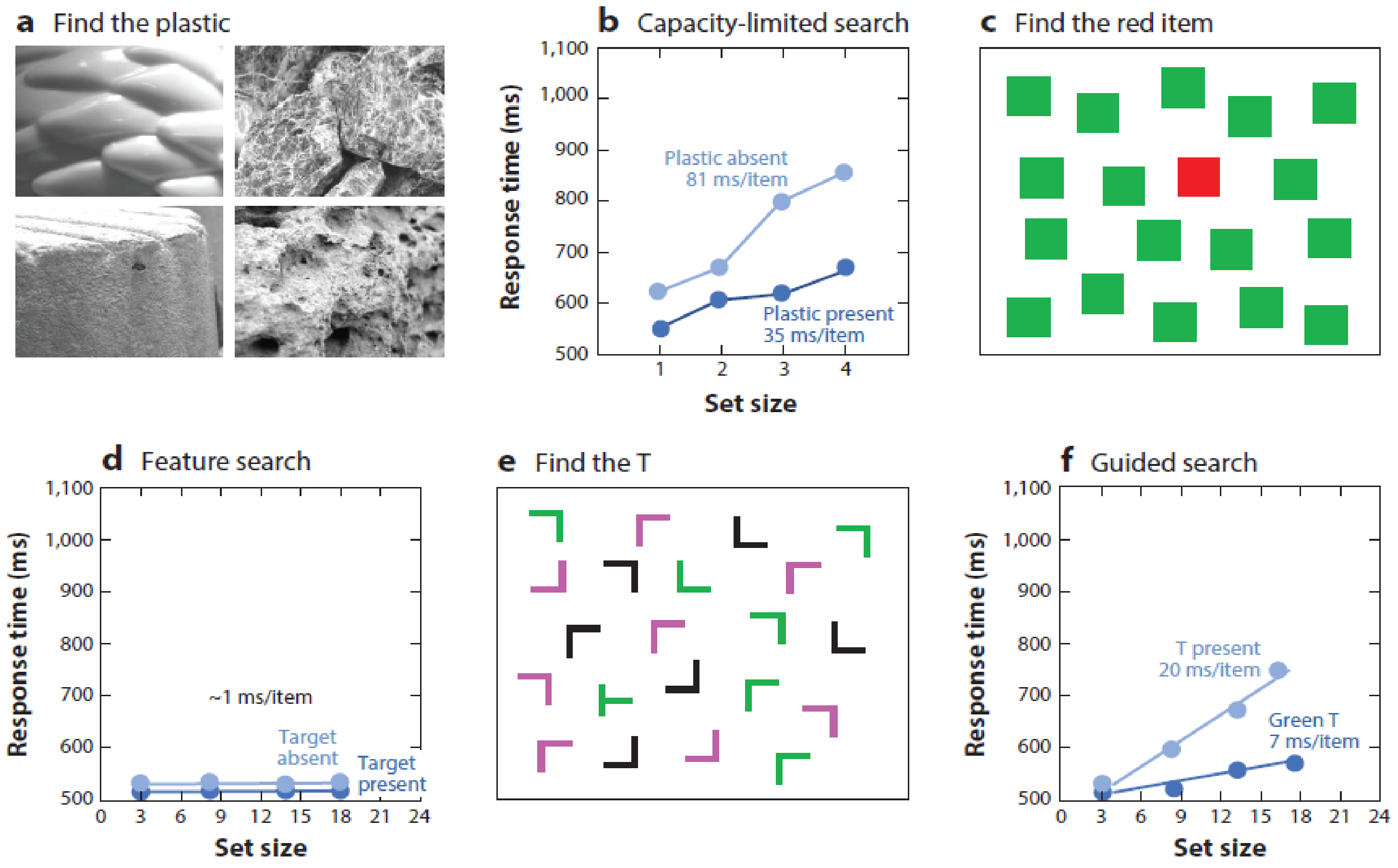


Figure 1

Classic visual search experiments. (*a*) Identifying plastic is easy even if defining plastic is hard. (*b*) Search for plastic requires an apparently serial search (data taken from Wolfe & Meyers 2010). (*c*) The red item pops out among green distractors. (*d*) In a search task, response time would be roughly independent of the number of green items in the display (data taken from Wolfe et al. 2010b). (*e*) Search for the T is more efficient if you know that the T is green. (*f*) This guidance by color is seen in the shallower slopes of the response time × set size function for search for green Ts, as opposed to search for any T (hypothetical data; see Egeth et al. 1984).

- Serial or parallel?
- Voluntary saccadic eye movements occur at a rate of about 3–4 per second, translating to RT × set size slopes of 250–333 ms/item for target-absent trials and approximately half that for target-present trials.
- In most real-world search tasks, initial parallel processing of the search display can be used to guide subsequent deployment of attention (1e). Searching for green T is easier because set size is smaller => slope is 1/3 the size of the slope when colour unknown.
- More guidance -> more efficient search

Guiding features

- Limited set of features: colour, size, motion, lighting direction, axis of rotation, etc.
- Feature guidance is derived from the basic sensory input that gives rise to the perception of attributes like color or orientation.
- guidance is a specific abstraction from that input (a 0° target will not pop out in a display of 5° distractors)
- guiding signals appear to be categorical (easier to search for oriented lines if they are uniquely steep or shallow (Wolfe et al. 1992))
- guidance is often usefully described as a relationship between targets and distractors, rather than as an absolute target value (orange target could be defined as the reddest item in one context but might be searched for as the yellowest item in another context)
- Guidance by shape is most complex and not researched enough (consists of other basic features, but are they processed independently or do they provide a single coarse representation?)

Feature guidance isn’t enough

- Classic feature guidance is guidance to locations or objects that appear to have target features.
- Scene guidance is guidance to locations that are more likely to contain a target, regardless of whether that target and/or its features are present.

Figure 2

Scene guidance: where observers look when told to look for a pillow. Figure reproduced with permission from Sage Boettcher and Melissa Vo, copyright 2019.

‘Meaning maps’

- Henderson & Hayes (2017, 2018)
- Similar to I&K salience maps, their meaning maps show how the meaningfulness of a scene varies over the scene.
- ‘scenes, like language, have a grammar.’ That grammar is concerned with scene semantics (Is an object meaningful in this location in this scene?) and scene syntax (Where is it structurally plausible for an object to be? (toasters do not float)) (Vo & Henderson 2009).

Scene guidance + feature guidance

Figure 3

The interaction of scene and feature. The observer is asked to look for people.

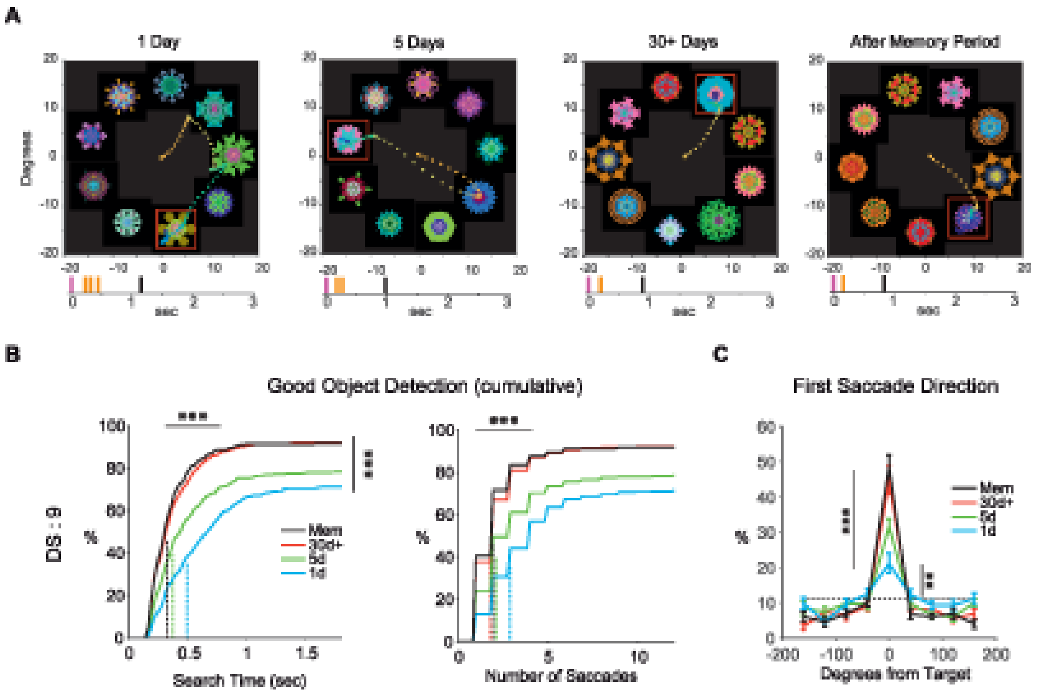
- if you look for people, you will quickly find the man in the middle of the image.
- You may be slower to see the exact copy of his image, pasted on the path at the bottom. You guided your attention to objects or features that could plausibly be human, and the man at the bottom is the wrong size.
- He is only the wrong size, however, because the structure of the scene makes it so (Eckstein et al. 2017,Wolfe 2017).

Guidance by Search History

- Performance on one search is influenced by performance on previous searches.
- Maljkovic & Nakayama (1994): observers were faster if the preceding target was the same color as the current target. The result has been replicated and extended to multiple dimensions (Hillstrom 2000, Kristjansson 2006).
- *finding something on one trial appears to guide attention toward subsequent instances of that target.*
- Originally, *Wolfe et al.* (2003): priming was a form of top-down guidance. (Bottom-up guidance from the stimulus, top-down from the observer. Since priming was a form of memory, it was internal to the observer -> top down)
- *Awh et al.*(2012): There is top-down, volitional guidance to what you are looking for, which is distinct from the automatic effects of history that occur whether you are looking or not.
- *Theeuwes* (2018): much of what passes for top-down guidance is driven by priming by targets.
- *Ásgeirsson & Kristjánsson* (2019): the recent history of search guides subsequent search.

Guidance by Value

- Attention is guided to features that have been associated with reward (Anderson et al. 2011).
- Similar to priming in that effects of value can be quite automatic and can act against the interests of the observer (Hickey et al. 2010).
- Effects of value can be seen in monkeys (Ghazizadeh et al. 2016):



Effects of the repeated object-reward association on the detection of Good objects (DS: 9). (A) Example search performance of monkey R after different training amounts and memory period. Eye position is shown by time-dependent color-coded dots (2/ms dot, from orange to blue). Red square indicates Good object (not shown to the monkey). Tick marks at bottom show the timings of saccades (orange) and reward (black) relative to display onset (purple). (B) Search time and number of saccades (cumulative) for detecting Good object, shown separately for different training amounts (left and right, respectively). Dotted lines: average of median across search sessions. Detection of Good objects shows significant increase and median search time and saccade number shows significant decrease by longer reward training. (C) Distribution of the first saccade directions toward nine equally spaced objects relative to Good object for different training amounts. Dotted line: chance level. First saccade toward Good object was already higher than chance even after 1-day training and significantly increased by longer reward training. Data in (B) and (C) are from all four monkeys. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Priority Maps

- Selective attention has a single focus -> the multiple sources of guidance (the 5 factors) need to be combined with some final common path.
- This is often imagined as a **priority map** consisting of some weighted average of all forms of guidance (Fecteau & Munoz 2006, Serences & Yantis 2006). (*The term salience or saliency map should be reserved for bottom-up, stimulus-driven guidance*)
- I&K: Information is pooled, and when it is time to deploy attention, a WTA process directs attention to the point of highest activation in the map.
- Chan & Hayward (2009): it is possible to respond based on activity in an orientation map or a color map without the need to combine those signals into a priority map.
- Buetti et al. (2019): combination of signals does occur. When the target was defined by both color and shape, RTs were faster than responses to either feature alone. This indicates that *combining features makes a bigger signal, driving faster responding.*
- Yantis 1993: Attentional capture suggests that activity in the priority map is unlabeled, in the sense that attention can be captured by the wrong signal if that signal is incorporated into the priority map.

Parallel + serial in guided search. Overt + Covert

- In the real world overt eye movements and covert deployments of attention should be seen as participating in a complex, interactive dance during search.
- Original GS: eyes move to a new fixation every few hundred ms, and that covert attention is then serially deployed to 5–10 items before the eyes move elsewhere.
- Functional visual field (FVF): attention might move to a location, permitting parallel processing of all items within a FVF surrounding that point of fixation. (the difference in efficiency of the search for a T and for a green T is not a difference in guidance but a difference in the size of the FVF.)

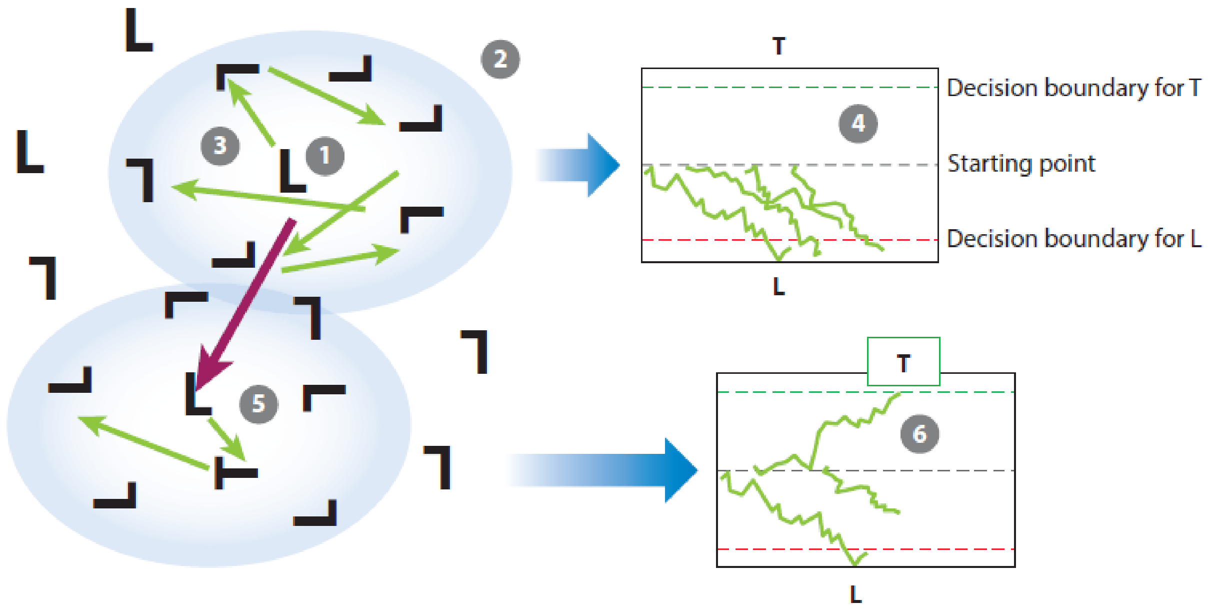


Figure 4

The dance of overt eye movements and covert attention. ① The observer fixates at some point. ② Within a functional visual field, shown by the dotted line, ③ they make covert deployments of attention (*green arrows*). ④ Each attentional selection starts a diffusion process, designed to identify the selected item as a T or an L. In this case, all processes reach the L decision bound. ⑤ Having failed to find a target, the eyes are redeployed to a new spot (*purple arrow*), surrounded by a new functional visual field. Again, covert deployments of attention are made. ⑥ This time, one of those items accumulates the information required to reach the decision boundary identifying it as a T.

Search templates & WM

- Visual search implies having a search template. (search template (Rajsic et al. 2017), target template (Bravo & Farid 2016), memory template (Kristjánsson et al. 2018), and attentional template (Yu & Geng 2019).)
- Guiding template (WM) - search template for a coarse WM representation that guides search
- and Target template (LTM)- LTM representation of the precise goal of the search.

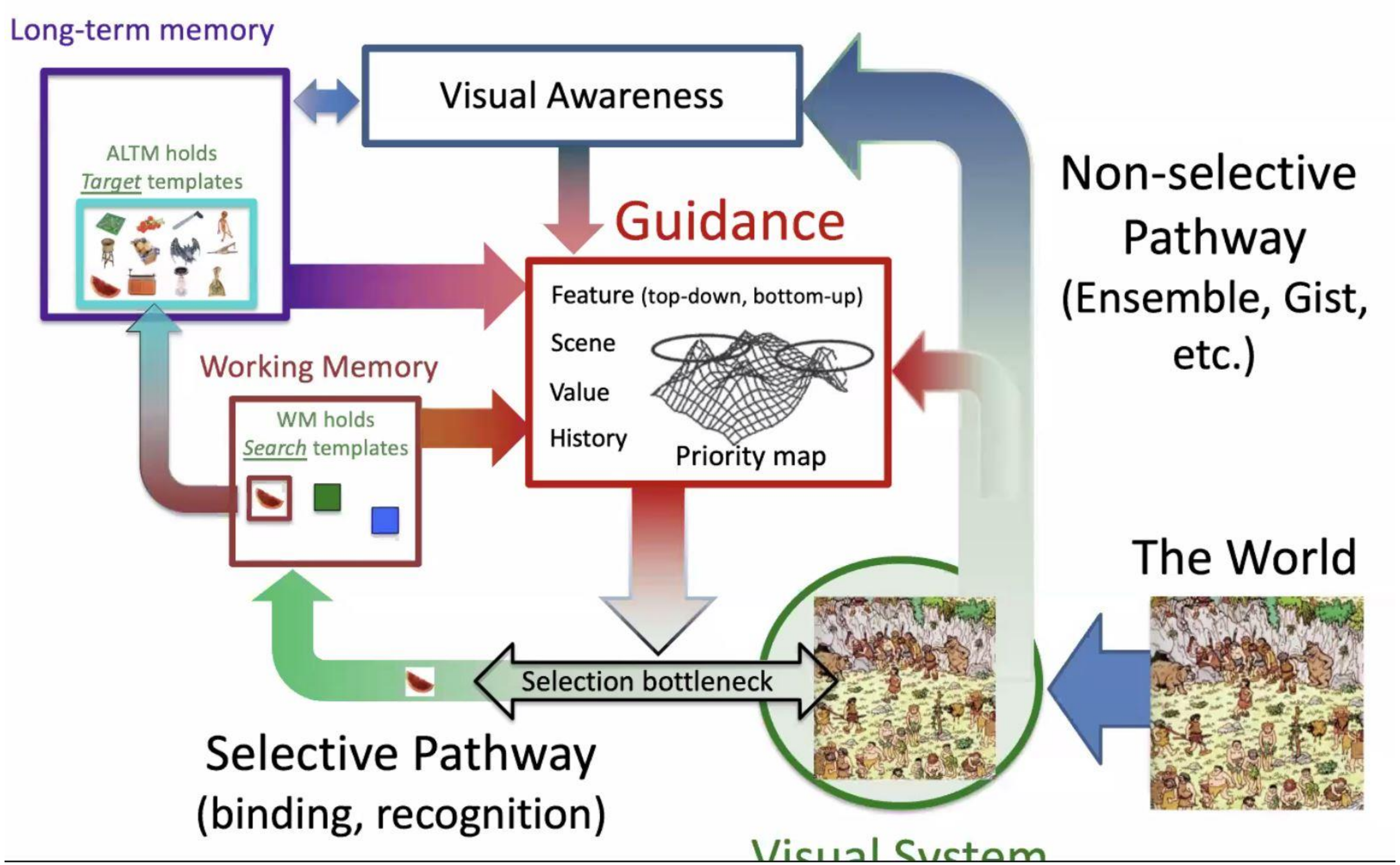


Image from J. Wolfe's online seminar that took place on Feb 4th (https://www.world-wide.org/seminar/2037/)

Search termination

When is it time to quit the current search task?

- FIT: 2-alt forced choice (target present or not)
- GS2: target found or when all distractors having activation or priority above some threshold had been examined.

rejected distractors or regions are marked in some way:

- IOR (Klein & MacInnes, 1999)
- visual search has no memory for rejected distractors (Horowitz & Wolfe)
- Static vs Dynamic (Shi et al):
  - Static: “Shall I quit, because I have looked at more or less everything that needs to be looked at?”
  - Dynamic: “Have I searched long enough that it is unlikely I would have missed the target?”
- Evidence of some limited memory during search (Dickinson & Zelinsky 2005, McCarley et al. 2003).

- Search termination becomes even more complicated when single-target search expands into foraging and/or beyond lab space into the open world
- Maybe quitting mechs include feedback about errors (Chun & Wolfe 1996), target prevalence (Wolfe & Van Wert 2010), and developing expertise (Brams et al. 2019).

- observers probably adjust quitting thresholds within a single search based on an evolving assessment of the current stimulus (**competitive GS model of Moran et al**. (2013)):

$$P(quit) = Wquit/(Wquit + \Sigma(stimuli))$$

where $Wquit$ is a quitting signal that grows on each step, and $\Sigma(stimuli)$ is some assessment of the likelihood that a target could be present. Early in a search, $P(quit)$ will be dominated by stimulus. As time goes on, $Wquit$ will push $P(quit)$ toward 1.0.

- One possibility, captured by the Moran et al. (2013) model, is that a larger stimulus signal on target-present trials keeps observers searching longer when there is, in fact, a target to be found.
- This topic deserves further research.

Psychonomic Bulletin & Review

VOLUME 20, NUMBER 4 ■ AUGUST 2013

PB&R

EDITOR

Cathleen M. Moore, University of Iowa

ASSOCIATE EDITORS

A PSYCHONOMIC SOCIETY PUBLICATION

Springer

Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-020-01859-9

THEORETICAL REVIEW

Guided Search 6.0: An updated model of visual search

Jeremy M. Wolfe [1,2]

Accepted: 4 December 2020