

The impending demise of the item in visual search

Johan Hulleman

Division of Neuroscience and Experimental Psychology, School of Biological Sciences, The University of Manchester, Manchester M13 9PL, United Kingdom.
johan.hulleman@manchester.ac.uk
<https://www.research.manchester.ac.uk/portal/johan.hulleman.html>

Christian N. L. Olivers

Department of Experimental and Applied Psychology, Institute for Brain & Behaviour Amsterdam, VU University, 1081 BT Amsterdam, The Netherlands.
c.n.l.olivers@vu.nl
<http://www.vupsy.nl/staff-members/christian-olivers/>

Abstract: The way the cognitive system scans the visual environment for relevant information – *visual search* in short – has been a long-standing central topic in vision science. From its inception as a research topic, and despite a number of promising alternative perspectives, the study of visual search has been governed by the assumption that a search proceeds on the basis of individual items (whether processed in parallel or not). This has led to the additional assumptions that shallow search slopes (at most a few tens of milliseconds per item for target-present trials) are most informative about the underlying process, and that eye movements are an epiphenomenon that can be safely ignored. We argue that the evidence now overwhelmingly favours an approach that takes fixations, not individual items, as its central unit. Within fixations, items are processed in parallel, and the functional field of view determines how many fixations are needed. In this type of theoretical framework, there is a direct connection between target discrimination difficulty, fixations, and reaction time (RT) measures. It therefore promises a more fundamental understanding of visual search by offering a unified account of both eye movement and manual response behaviour across the entire range of observed search efficiency, and provides new directions for research. A high-level conceptual simulation with just one free and four fixed parameters shows the viability of this approach.

Keywords: attention; eye movements; features; fixations; functional field of view; oculomotor control; visual search; visual selection

1. Introduction

Whether we are trying to find a friend amongst disembarking passengers or looking for a street name to establish our whereabouts, searching for targets is a ubiquitous part of our lives, and it involves fundamental cognitive mechanisms of perception, attention, and memory. Therefore, determining how we scan the visual environment for relevant information is a fundamental goal of vision science.

Visual search behaviour has been studied for a long time. For example, the very first issue of the *Quarterly Journal of Experimental Psychology* contained a paper on it (Mackworth 1948), and Neisser wrote about finding a face in the crowd for *Scientific American* back in 1964. But the two most seminal years in the field of visual search probably lie in the 1980s. At the start of that decade, Treisman and Gelade (1980) published their classic Feature Integration Theory (FIT). At the end, Wolfe et al. (1989), as well as Duncan and Humphreys (1989) proposed their very influential alternatives, Guided Search (GS) and Attentional Engagement Theory (AET). These contributions made visual search a burgeoning research area. In fact, they have been so successful that a recent review of visual search, published almost 25 years later, still listed FIT,

GS, and AET as the leading theories (Chan & Hayward 2013). However, although these dominant theoretical frameworks have inspired great advances in the study of visual attention, in our opinion, further progress is hindered by what appears to be an implicit yet central assumption, namely that the primary unit of selection in visual search is *the individual item*.

In the lab, the typical search experiment involves a single known target, which can range from a simple geometrical shape to a more complex alphanumeric character or an everyday object. Participants are usually instructed to determine its presence amongst a varying number of distractor items, although there are variants of the task in which the target is always present and observers make a decision on some orthogonally varied property (e.g., the identity of a letter inside of a target that is defined by colour). The effect of the number of distractor items on RT – the slope of the search function – is an important measure, because it indicates how efficiently observers detect the target. Although theories of visual search broadly recognize that there is a large amount of parallel processing within the visual field, this has had surprisingly little impact on what has been assumed to be the core

process, namely the selection of individual items that are either rejected as distractors or recognized as a target. There are a number of promising alternative perspectives that ground the search process in eye fixations rather than covert selections of individual items. We argue that these approaches, when unified, provide a more comprehensive framework for explaining the oculomotor and manual response dimensions of visual search behaviour. The goal of this paper is to provide this unification.

2. Setting the stage: Feature Integration Theory and its assumptions

FIT was never intended as a theory of visual search proper, but rather used the visual search paradigm to test its predictions about the way early sensory processing produces object representations. Nevertheless, it is difficult to overestimate its influence on the formulation of the visual search problem. The fundamental distinction between search with flat slopes, where the time taken to find the target is independent of the number of distractors (e.g., / amongst |), and search with steeper slopes, where search time increases with set size (e.g., red / amongst green / and red |), had been made before (e.g., Jonides & Gleitman 1972). But Treisman and Gelade's (1980) FIT provided an attractive explanation. In its original version, visual features (e.g., colour, orientation, motion) are pre-attentively registered in parallel in separate feature maps. So, whenever the target differs from the distractors by a single feature (e.g., red amongst green), search time is independent of set size. Target presence is simply established by inspecting activity in the relevant feature map. Identifying a target that is a conjunction of features (e.g., red | amongst green | and red /), however, requires serially applied attention to bind the features together, using a map that contains the item locations. Consequently, whenever the target is defined by a combination of features, RTs increase with set size. Thus, FIT explained the quantitative difference between single feature and conjunction search slopes as a qualitative difference between parallel, "map"-based search and serial, "item"-based

search. As we will see later, this qualitative distinction prompted an enduring empirical focus on the shallower end of the search slope spectrum as the most informative about the fundamental mechanisms of visual search. After all, somewhere between 0 ms/item and around 25 ms/item (for target-present trials) the transition to item search occurs. Consequently, search beyond this range has been considered to have little additional theoretical value.

FIT opened up an abundance of research questions. It predicted binding errors, where features are combined incorrectly (e.g., Treisman & Schmidt 1982). It also inspired a taxonomy of basic features, by providing the diagnostic of flat search slopes (see Wolfe & Horowitz 2004, for an overview). And importantly, because of its fundamental distinction between parallel feature search and serial conjunction search, FIT encouraged other researchers to challenge the core of the theory by finding conjunctions of features that nevertheless yielded flat search slopes. Success in this endeavour (e.g., Nakayama & Silverman 1986; McLeod et al. 1988; Wolfe et al. 1989) gave rise to new models (Duncan & Humphreys 1989; Wolfe et al. 1989) and to adaptations of FIT (Treisman & Sato 1990; Treisman 1991).

3. Popular alternative theories: Guided Search, Attentional Engagement Theory, and Signal Detection approaches

3.1. Guided Search

Guided Search, the hitherto most successful model, was conceived to challenge FIT's fundamental distinction between parallel feature and serial conjunction search. Wolfe et al. (1989) adapted FIT such that information from the feature maps guides attention towards conjunctions as well. Across several updates (Wolfe 1994; 2007; Wolfe & Gancarz 1996) the basic principle has remained unchanged: Guided Search combines signals from different feature maps into a single activation map via broadly tuned ("categorical") channels (e.g., "red," "green," "vertical," "horizontal"). The activation map holds the locations of the individual items, and attention is guided towards the location with the highest activation. If it contains the target, a target-present response follows. However, because of inherent noise, it may contain a distractor. In that case, attention is guided to the location with the next-highest activation. This continues until the target is found or search is terminated with a target-absent response.

Top-down weighting or filtering of the channels improves search efficiency. For example, for a green-horizontal target and distractors that are red-horizontal and green-vertical, output from the green and horizontal channels is selected. Because the target receives activation from two channels, while distractors receive enhancement from only one, attention can be more efficiently guided towards conjunction targets, allowing for relatively flat search slopes. Furthermore, top-down weighing of specific features explains why people often search through or ignore subsets of items (e.g., Kaptein et al. 1995; Watson & Humphreys 1997). Accordingly, there is no fundamental distinction between feature search and conjunction search, making both essentially item-based.

JOHAN HULLEMAN is a lecturer in the School of Biological Sciences at The University of Manchester. He received his Ph.D. from the University of Nijmegen. His main interest is in visual perception, visual attention, and visual search. He has (co-)authored over 30 publications on these topics.

CHRIS OLIVERS received his Ph.D. in 2001 from the University of Birmingham, United Kingdom. He is now Full Professor of Visual Cognition at the Institute of Brain and Behaviour, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam. His research focuses on visual perception, attention, multisensory processing, working memory, and cognitive control. He has (co-)authored more than 100 publications on these topics and has received awards from the British Psychological Society and the American Psychological Association.

The latest version of Guided Search (4.0; Wolfe 2007) differs from its best-known predecessor Guided Search 2.0 (Wolfe 1994) in the way individual items are selected. In version 2.0, items were selected and processed individually in a purely serial fashion at a rate of 50 ms per item. In version 4.0, items are also selected individually and at a similar rate (20–40 ms per item), but they now enter a processing stream that itself takes between 150–300 ms to establish whether an individual item is the target. This component was added to account for findings from attentional dwell time studies, which suggest that items need this amount of time to be processed (Duncan et al. 1994; Theeuwes et al. 2004). The stream has a capacity of four items. Guided Search 4.0 is therefore no longer a purely serial model, but a serial–parallel hybrid, and is often referred to as the *car wash model*. Yet, even in version 4.0, the individual item remains at the heart of the search process. Although multiple potential targets are processed simultaneously, these candidates are still delivered one-by-one to the car wash. And despite the disavowal of the qualitative distinction between flat and steeper search slopes, the shallow end of the search slope spectrum continues its important theoretical role, because that is where visual properties that support top-down guidance are separated from those that do not, allowing conferral of the theoretically important concept of “feature status” on the former.

3.2. Attentional Engagement Theory

Another challenge to FIT came from Duncan and Humphreys (1989), who also criticized the dichotomy between parallel and serial search, but on different grounds. In what later was called AET (Duncan & Humphreys 1992), they proposed a continuous search *surface*, where the combination of target–distractor and distractor–distractor similarity determines a range of search slopes. When distractors resemble the target, search times increase. When all distractors resemble each other, search times decrease. Hence, search must take the relationship between multiple items into account, rather than just the identity of single items.

It is fair to say that Duncan and Humphreys never envisaged a theory purely based on individual items. Instead, they proposed that search operates on “structural units” – segments in a hierarchically organized representation of the visual input that may be defined at various levels (from individual items to scene structures – see also Nakayama & Martini 2011). These structural units compete for access to visual short term working memory (VSTM). The better the match with the target template, the higher the probability that a structural unit enters VSTM; the better the match with a distractor template, the lower this probability becomes. Changes in the selection probability of a structural unit spread in parallel to similar structural units throughout the display.

Yet, although AET was set up as a theory about structural units, its subsequent application to visual search has essentially been item-based. As Duncan and Humphreys (1989, p. 446) state: “In an artificial search display it may seem reasonable to limit consideration to the few stimulus elements that are presented by the experimenter, but in a realistic, natural image the problem is more complex.” In their account of visual search data, they continue: “[T]here is

the problem of classifying each single element in a display as target or non-target. In [AET] this involves matching each element against a template of possible targets” (p. 447). An item-based approach is also notable in SERR (Humphreys & Müller 1993), a computational implementation of AET. Here, the individual identity of items (e.g., a particularly oriented T or L) is compared against templates specifying the identity of individual targets and distractors, although items can be strongly grouped if they are of the same identity. So a T is rapidly detected among Ls because the grouped Ls provide strong evidence for the presence of an L and will cause a match with the L-template. This is then followed by inhibition of all Ls, applied via their individual locations, leaving the T as the last uninhibited item. Consequently, the group process is still based on the identities and locations of individual items. Furthermore, the associated empirical work focused on relatively shallow search slopes. Of course, in principle, AET can be applied to structural units other than individual items or to more difficult search. So far, however, AET has not been extended beyond easier, item-based search.

3.3. Approaches based on Signal Detection Theory (SDT)

SDT approaches to visual search (e.g., Eckstein et al. 2000; Palmer et al. 2000; Verghese 2001) form a different class of theory and are explicitly formulated as a rejection of the two-stage architectures of FIT and Guided Search. Instead, SDT approaches assume a single, parallel stage during which the target and distractor items evoke noisy individual internal representations, with the target’s representation scoring higher along the relevant feature dimension. Importantly, because of neural noise, there will be an overlap in the distribution of these individual internal representations. The more similar target and distractors are, the larger this overlap. Target-absent and target-present responses are based on a decision rule. A popular choice is the MAX-rule, where the decision is based on the single item with the largest feature score. The larger the number of distractors, the higher the probability that one of them evokes an internal representation that is target-like. Therefore, evidence for target presence decreases with set size.

Their fundamental opposition to FIT and Guided Search notwithstanding, SDT approaches so far have shared their item-based nature. Even though displays are processed in parallel, decisions are still based on the internal representations evoked by individual items. Moreover, in conjunction searches, the location of the individual items is used to combine the representations on different feature dimensions. Finally, as the main aim of SDT theories was to provide an alternative explanation for flat versus steeper slopes, they too have focused on the shallow end of the search spectrum.

4. The problem: Why items as the conceptual unit hinder more than help in understanding visual search

We hold that the focus on the item as the core unit of visual search is rather problematic for a number of reasons.

4.1. It ignores other ways of doing visual search

Item-based approaches limit the real-world applicability of results from the lab. In that sense, the adoption of the item as conceptual unit may have had an effect on the type of stimuli used as well: Item-based models make item-based predictions that are tested with item-based displays. Yet, although radiologists and security-screeners undoubtedly perform visual search tasks, it is not immediately clear how many “items” mammograms or airport security X-rays contain. Neider and Zelinsky (2008) argued convincingly that it is impossible to objectively define set size in real-world scenes. Similarly, using the individual item as conceptual unit requires a distinction between texture displays (with many items, or all items forming a coherent surface) and search displays (with fewer items). Although Wolfe (1992) reported a dissociation between groups of stimuli that allow texture segmentation and individual items that allow efficient search, it remains unclear how many items are needed before a search display becomes a texture. We are not saying that proponents of item-based models ignore real world searches. On the contrary, the two main authors of the original Guided Search model, for example, regularly publish admirable work on search in real world scenes (Vö & Wolfe 2012; Wolfe et al. 2011b), medical imaging (Donnelly et al. 2006; Drew et al. 2013a; 2013b; Evans et al. 2013a), and luggage-screening (Godwin et al. 2010; Menneer et al. 2007; Wolfe et al. 2005; 2013). However, as Wolfe et al. (2011b) pointed out, classic item-based models generally fail under these circumstances. To account for scene-based search, yet also preserve the item-based structure of Guided Search, Wolfe (2007; Wolfe et al. 2011b) assumes a pathway for scene processing that is separate from item-based processing.

Conceptualizing search as being based on selecting individual items limits thinking about alternative ways to complete the task. The item as conceptual unit has made it tempting to view search as a process where items are compared against a template specifying the individual target item (e.g., Bundesen et al. 2005; Humphreys & Müller 1993; Wolfe 1994; Zelinsky 2008) and possibly also other types of target-defining information. This item-based template-matching then provides an underlying rationale for reporting visual search experiments in terms of RTs as a function of set size, where the slope measures the additional cost of having to compare an extra item to the template. However, item-based approaches encounter the problem that search slope estimates of individual item processing (typically 25–50 ms/item) are much lower than estimates of attentional dwell time from other paradigms, which have reported item-processing times of 200–300 ms (Duncan et al. 1994; Theeuwes et al. 2004). This is why Moore and Wolfe (2001) proposed the car wash model: Search slopes measure the rate at which individual items are entered into a processing stream, rather than processing duration itself, in the same way that the time between two cars entering a car wash can be shorter than the time it takes to wash an individual car. But this model is only necessary if one conceptualizes visual search as the problem of linking one item-based process (a fast serial search of 20 to 40 items per second) to another (a slow bottleneck of about 4 items per second).

In many visual search experiments though, the task is to decide whether the display contains a target – not whether

any specific item is a target or a distractor. Conceptualizing the search process as a sequence of item-based present/absent decisions is potentially misleading, because checking whether a particular item is the target is not the only way to complete the task. For instance, looking for a difference signal between the target and its surrounding distractors might work too. This possibility was first recognized for simple feature searches, where “singleton detection mode” (search for any difference) has been distinguished from “feature search mode” (search for a specific feature; Bacon & Egeth 1994; or feature relationship, Becker 2010). Another promising alternative formulation was given by Rosenholtz et al. (2012a), who proposed that observers decide whether a particular fixated patch of the search display contains the target on the basis of pooled summary statistics computed across that patch. Evidence against single item approaches comes from a computational model of Najemnik and Geisler (2008). They argued that human eye movement patterns during visual search are better explained by a model that fixates areas of the screen that maximize information about target-presence, than by a model that fixates the item most likely to be the target (see also Young & Hulleman 2013). Likewise, Pomplun and colleagues (Pomplun et al. 2003; Pomplun 2007) reported that fixation patterns not only depend on the presence of particular relevant or irrelevant features, but also on the specific local ratios and spatial lay-outs of these features – that is, local statistics. They too found fixations often to be off-items. This behaviour was successfully captured in a model that assumes *area* activation rather than individual item activation. Thus, decisions about target-presence could very well be framed at the level of group-statistics of items, rather than at the level of individual items. Item-by-item search may actually be the exception. When search does proceed item-by-item, as demonstrated by fixations on each individual object, performance becomes very poor (Hulleman 2010; Young & Hulleman 2013), with extremely steep search slopes and miss rates exceeding 20%. Performance in standard laboratory tasks is typically much better, suggesting less effort is involved than predicted by item-based theories. The idea of items processed in spatial clusters is not new. Pashler (1987) already proposed search through clumps of items, and arrived at a fixed clump size of 8 items, with 75 ms for every between-clump switch, and a ± 15 ms/item slope for within-clump search, although he also argued that it may vary with different types of a search.

One might argue that not all search tasks can be based on global statistics because some really do require the individual item. For example, tasks may involve a response to the precise location of the target, or to a relatively difficult to distinguish property that is varied orthogonally to the target-defining feature. This latter type of task is often known as compound search (Duncan 1985), and may partly involve processes that differ from a present/absent task (e.g., Olivers & Meeter 2006). However, the fact that the individual target item is required at the end does not mean that the preceding search process is also item-based. Search could be conceived as consisting of multiple steps where statistical signals are used to select the rough area containing the target, more precise signals are then used to exactly locate it, finally followed by even more precise extraction of the response feature. The first steps are likely to be very similar across search tasks, while the

later steps are likely to differ depending on what exactly is required from the target (see Töllner et al. 2012b, for direct evidence).

4.2. It overestimates the role of individual item locations

A main reason why individual items play such an important role in visual search theories is that their locations are necessary for effective feature binding (in FIT, but also in Eckstein et al. 2000; Itti & Koch 2000; and Wolfe 2007), or for collecting the features used in guiding attention and template-matching. Moreover, individual item locations are needed to inhibit previously inspected distractors. Yet, visual search is very robust against substantial displacement of items, at least for present/absent tasks. Horowitz and Wolfe (1998) reported largely intact search performance when items are randomly shuffled around the display every 100 ms. Furthermore, Hulleman (2009; 2010) reported that search for a moving T amongst moving Ls remained comparable to search in static displays, even when dozens of items were moving smoothly in random directions at velocities of up to 10.8 deg/s (and tracking of all individual items is virtually impossible; Pylyshyn & Storm 1988). Even occupying observers with an additional working memory task hardly affects search through such random motion displays (Hulleman & Olivers 2014). These results suggest that the exact location of individual items is less important than previously assumed by item-based accounts. Instead, they support the idea that present/absent decisions are based on parallel extraction of properties of *groups* of items within local areas; properties that are holistic or statistical in nature. An example is the pooling approach of Rosenholtz et al. (2012a) mentioned earlier. Here, summary statistics (for relative orientation, relative phase, correlations across scale, etc.) are computed across a patch of the search display. This means that the locations of individual items inside the patch are inherently less important than the location of the patch in the display. Item motion ceases to be a special case because individual location information is also discarded for static items. Finally, note that a pooling approach is less taxing on memory: no memory for individual items is needed, only for inspected areas.

4.3. It ignores a really difficult search

The influence of FIT's distinction between parallel feature-based search and serial item-based search has resulted in an unwarranted emphasis on the shallow end of the search slope spectrum. For example, in Wolfe's (1998b) analysis of more than 1 million search trials, 90% of the target-present slopes were below 40 ms/item. We suspect that more difficult search tasks are used only sparingly because of the FIT-derived idea that search becomes completely item-by-item once you have crossed the 25 ms/item barrier (T vs L; 2 vs 5; as most explicitly stated by Wolfe 2007). Once this item-by-item stage has been reached, there is little extra theoretical insight to be gained from even slower search, because any additional slowing cannot be due to the core search process.

Furthermore, it appears that slope differences at the shallow end are still given a qualitative interpretation; they are seen as *diagnostic* for visual properties that support top-down guidance and thus have "feature

status." For example, Wolfe (2007, p. 106) writes that a T is easily discriminable from an L, just like a \ is easily discriminable from |. Yet search for T is inefficient (25–50 ms/item), and search for \ is parallel (0–10 ms/item). Thus, within the Guided Search framework, the conclusion is that orientation guides attention, while T or L junctions do not, and therefore that somewhere between 10 and 25 ms/item there is an important transition. Note further that Guided Search thus explicitly dissociates *discriminability* from *feature guidance* (cf. Beck 1972; Beck & Ambler 1973): An easily discriminable visual property is not necessarily a guiding property. This is counterintuitive because one would expect that the visual system will use properties that it finds easily discriminable.

The focus on the shallow end of the search slope spectrum has led to an explanatory gap at the steep end. For example, search for T amongst Ls is considered a prototypical example of a task where differences between target and distractor are at an absolute minimum. Both consist of the same two lines and only the relative position of these lines determines whether an item is a target or a distractor. The associated slope values of 25 ms/item and 50 ms/item (target-present and target-absent, respectively) should therefore constitute an upper limit for the steepness of search slopes. However, Wolfe (1998b) reported searches with slopes much steeper than 25–50 ms/item, even up to 100–250 ms/item. This makes additional hypotheses necessary. For example, very slow search may be due to hard to discriminate objects (perhaps requiring serial extraction of features *within* an item, thus slowing down the car wash), or due to eye movements. Such discriminability and eye movement influences may indeed be fundamental to hard search, but, as we will argue later, the same factors may in fact explain all search slopes. That is, there is no need for additional hypotheses to explain steep slopes, but for a single hypothesis that explains all slopes.

The explanatory gap becomes even wider if one considers that if a slope distinction that suggests qualitative differences actually exists, it appears to occur at the high end of the slope spectrum, at values of around 100 ms/item or more. Up to a few tens of milliseconds per item, search is quite robust against item motion, but very slow search (of 140 ms/item) breaks down when items move (Hulleman 2009; 2010; Young & Hulleman 2013). In contrast, very slow search is robust in gaze-contingent displays where only the fixated item is unmasked, whereas easy to moderate search becomes much slower and more error-prone when the number of unmasked items is reduced (Young & Hulleman 2013). Easier and very hard search also differs in terms of RT distributions. Young and Hulleman (2013) found that for easy (0 ms/item) and intermediate search (± 20 ms/item target-present), the standard deviation of the RTs was larger for target-absent trials than for target-present trials (see also the searches of up to about 40 ms/item for target-present in Wolfe et al. 2010a). On the other hand, the pattern is reversed for hard search (± 140 ms/item target-present): Here, the standard deviation of the RTs is largest for target-present trials (Young & Hulleman 2013). Later we will explain what we believe to be the origin of this differential robustness to motion, differential robustness to visible area size and reversal in variability in RTs. The point for now is that the emphasis on the differences at the shallow end of the search slope spectrum has resulted in an underappreciation of the

similarities between searches at that end, as well as their differences with searches at the steep end.

4.4. It ignores the eye

Understandably, when trying to explain the difference between feature search and conjunction search, or between guided search and unguided search, researchers have had to control for eye movements as a possible confound. In this sense, eye movements have traditionally been considered a nuisance phenomenon, rather than a crucial component of the search process. Although Treisman and Gelade (1980) acknowledged the role serial fixations may play in search performance, Treisman (1982) claimed that “serial search [...] is centrally rather than peripherally determined; it represents successive fixations of attention rather than eye movements” (p. 205–206), and later iterations of FIT (e.g., Treisman & Sato 1990; Treisman 1991) no longer mention eye movements. Pashler (1987), when discussing the possibility that items are processed in clumps rather than individually, decided that this was “not due to eye movements, in any interesting sense” (p. 200). Wolfe (1998a) shared this view: “While interesting, eye movements are probably not the determining factor in visual searches of the sort discussed in this review—those with relatively large items spaced fairly widely to limit peripheral crowding effects” (p. 14). Similarly, the AET and SDT approaches also did not account for eye movements. Even though these opinions were expressed decades ago, they continue to reflect mainstream thinking in the field.

Many models of search (e.g., Itti & Koch 2000; Wolfe 2007) have equated eye movements with *covert* shifts of attention, in the sense that overt shifts, when executed, simply follow the covert shifts. Stated the other way around, covert visual search is like overt visual search, but without the eye movements. The fact that search can proceed without eye movements is used as an argument that search is *de facto* independent of eye movements (see Carrasco 2011, and Eimer 2015, for more recent iterations of this view). This does not mean that these researchers deny that eye movements exist—or influence search—rather, they do not assign eye movements a central, explanatory role in modelling search behaviour. The equating of overt to covert shifts is convenient, as it allows eye movements to be disregarded. Visual search becomes a homogeneous sequence of shifts of attention, with the entire display at its disposal, rather than an amalgamation of different viewing episodes, each with their own start and end point, and each with their own spatial distribution of information. As support, Wolfe (2007) cites studies showing that, with appropriately scaled items, search with and without eye movements is comparable (Klein & Farrell 1989; Zelinsky & Sheinberg 1997). Yet, because covert shifts are assumed to operate at a faster pace than overt shifts, additional assumptions are needed (Itti & Koch 2000; Wolfe 2007). As Wolfe (2007, p. 107) states, “the essential seriality of eye movements can point toward the need for a serial selection stage in guided search.”

We agree that search can occur without eye movements (if the display allows), and that attention can be directed covertly—something we will return to in the General Discussion. However, there are also clear differences

between eye movements and covert attentional shifts. The latter are limited by the physiology of the retina, whereas the former are used to surmount those limitations. Emphasising the similarity between eye movements and covert shifts of attention by suggesting that, with appropriately scaled items, searches with and without eye movements yield similar results, ignores the reverse argument, namely that this similarity might not hold in most other situations, where items are typically not appropriately scaled. Under free viewing conditions there is a strong positive correlation between number of fixations and both task difficulty and RT (e.g., Binello et al. 1995; Motter & Belky 1998a; Young & Hulleman 2013; Zelinsky & Sheinberg 1995; 1997). Moreover, even when search could proceed without eye movements, participants still prefer to make them (Findlay & Gilchrist 1998; Zelinsky & Sheinberg 1997). The dominant models of search so far do not account for this fact, because they start from the position that successful search can occur without eye movements.

As argued by others (Eckstein 2011; Findlay & Gilchrist 1998; 2001; 2005; Pomplun 2007; Rao et al. 2002; Zelinsky 1996; 2008), the findings listed above suggest that eye movements are a fundamental part of visual search, and that any model without them is necessarily incomplete. We believe that not accounting for eye movements is not simply an omission or a matter of taste, but the logical consequence of adopting the individual item as the conceptual unit in visual search, with further consequences for visual search theory. For example, when eye movements are in principle unnecessary, and simply interchangeable with covert shifts of attention, the increase in number of fixations with increasing search difficulty becomes an epiphenomenon, necessitating the formulation of additional hypotheses—for example, feature binding, differential guidance, or differential attentional dwell times. As we will argue instead, it is more straightforward to assume that search RTs are directly related to the number of fixations. Then all that needs explaining is why some searches yield more fixations than others.

Taking eye movements into account requires acknowledging why they are needed to begin with. Consequently, the assumption that the entire search display is processed with the same detail no longer holds (Eckstein 2011). This assumption has been crucial to one of the main arguments for item-based feature binding accounts, namely that distinctions made equally easily in foveal vision (T vs. L, / vs. \, 2 vs. 5) yield very dissimilar search slopes (e.g., Wolfe & Horowitz 2004; Wolfe 2007). In other words, the argument here is that perfectly discriminable items nevertheless do not guide attention but instead lead to serial search—hence, feature binding implies item-based processing. However, this ignores the differential drop-off in identification rate for these stimuli across the retina (e.g., He et al. 1996). Whenever a search display cannot be foveated in its entirety, the relevant question becomes how far into the periphery target detections are possible. The further into the periphery such detections can be made, the fewer eye movements are needed, and the faster search will be. Clear eccentricity effects on visual search RTs have been reported (e.g., Carrasco et al. 1995; Motter & Belky 1998b; Scialfa & Joffe 1998). But retinal resolution also affects eye movements themselves. Young and Hulleman (2013) showed that the distance between fixation location and nearest item depends on

task difficulty. The easier the discrimination between target and distractor, the larger this distance was, and the fewer fixations were made.

We already referred to a number of very promising models that have taken the eye (either its movement or its retinal resolution) into account (Geisler & Chou 1995; Najemnik & Geisler 2008; Pomplun 2007; Pomplun et al. 2003; Rosenholtz et al. 2012a). Perhaps the most important model in this respect is TAM (Target Acquisition Model; Zelinsky 2008; Zelinsky et al. 2013). TAM is a pixel-based approach that was explicitly developed to model eye movements, rather than RTs. It has been very successful in explaining eye movement patterns in visual search, including fixation of empty regions to maximize population based information (Zelinsky 2012; Zelinsky et al. 2013). However, until now, all of these models have been models of fixations. They do not model slopes of manual RT functions or RT distributions, although it should not be too difficult to extend them and accommodate those measures (see Zelinsky & Sheinberg 1995 for an early proposal). Furthermore, in contrast to item-based models, fixation-based models have often focused on the *difficult* end of the search spectrum, using displays with targets that are very similar to the background or distractors. This is probably no coincidence, because it is these types of searches that are guaranteed to generate eye movements. In the next section, we will present a general framework intended as a bridge between eye movements and manual responses in search, across a range of search difficulties.

5. The solution: Towards fixation-based, rather than item-based search

So far, the main quest of visual search theories has been to account for the more central perceptual limitations affecting the search process, from feature binding to top-down guidance, from covert selection to inhibition of items, and from staggered serial (car wash) processes to post-selection bottlenecks. These limitations have been expressed as limitations of selecting and processing individual items. We agree that such central limitations on visual selection are important. However, the evidence reviewed suggests that the emphasis on individual items is becoming counterproductive, because (1) it obscures other theoretical possibilities that may be at least equally likely (e.g., using population-based signals), (2) it ignores earlier influences on the visual selection process that, because of ingrained physiological and other processing limitations, can be expected to have at least as profound an influence on visual selection as any central limitations, and (3) it has focused the research effort on easier search tasks to the detriment of further theoretical gains that harder search tasks could provide.

We believe that all components are in place for an overarching framework of visual search. One strand of the literature has provided models for RTs, while another strand has provided models for fixation behaviour. Although there have been fruitful attempts to link them (Geisler & Chou 1995; Zelinsky & Sheinberg 1995), these two strands appear to have grown further apart since. Making a link is not just a matter of combining the two strands. One type of model denies a pivotal explanatory role for eye movements in search, while the other type considers

them crucial. Thus, any overarching conceptual framework will require a fundamental, principled choice. We choose a framework that favours *fixations*, rather than individual *items*, as the conceptual unit of visual search. This has several advantages: Adopting fixations as the conceptual unit allows all kinds of displays into the visual search fold, including real world scenes and X-rays, rather than only those with clearly defined items. It also obviates the distinction between textures and search displays. A corollary of emphasizing the role of fixations in visual search is that retinal physiology becomes more important. This seems appropriate, because the maximum distance into the periphery where targets can be detected appears to be a major determinant of search times. Finally, a fixation-based framework allows for a much wider range of search slopes to be encompassed than the 0–50 ms/item on which the literature has typically focused. At the same time, adopting fixations as the unit of visual search does not negate the possibility of covert shifts of attention – something to which we will return in the General Discussion.

5.1. Functional Viewing Field

Central to the proposed framework is the Functional Viewing Field. As others have pointed out (see Eckstein 2011 for a review), retinal constraints are not the only limits on peripheral vision: Competition between representations occurs at many levels beyond the retina. For example, there are limits on attentional selection beyond that expected on the basis of visual acuity (Intriligator & Cavanagh 2001). There are also well-known effects of crowding and masking, where a stimulus – including simple features – that is perfectly recognizable on its own severely suffers when surrounded by other stimuli (Bouma 1970; Levi 2008; Neri and Levi 2006; Pelli et al. 2004; Pöder 2008; Pöder & Wagemans 2007). Even when limits to retinal and attentional resolution are taken into account, there remains a general bias to attend more to central items (Wolfe et al. 1998). Wolfe et al. (1998) argued that attention may follow the physiological constraints, such that areas of the retina that deliver the most information (i.e., the fovea) receive most attention. Therefore, observers may not always make eye movements out of bare necessity, but also out of efficiency or convenience. Furthermore, Belopolsky and Theeuwes (2010) have argued for a flexible “attentional window.” They reasoned that very easy search allows for a broad, more peripheral window, whereas hard search calls for a narrower, more foveal window.

The combination of peripheral constraints on perceptual and attentional resolution creates what has since the 1970s become known as the *functional viewing field*, FVF (Sanders 1970), the area of visual conspicuity (Engel 1971), visual span (Jacobs 1986; O'Regan et al. 1983), or *useful field of view*, UFOV (Ball et al. 1988). We will use FVF here and define it as the area of the visual field around fixation from which a signal can be expected to be detected given sensory and attentional constraints. Importantly, the FVF is not fixed but changes with the discriminability of the target. The less discriminable the target, the smaller the FVF, and the more fixations are needed to find the target. Hence, targets that are difficult to discriminate lead to longer search times. This even holds for search *without* eye movements, because targets that are less distinguishable from

distractors will suffer relatively more from additional distractors, especially in the periphery where discriminability will be lowest.

Direct support for the idea that FVF size distinguishes easy from hard searches comes from Young and Hulleman (2013), who masked peripheral information in a gaze-contingent design. The size of the unmasked region around fixation was varied from relatively large (about 10 degrees radius), to medium (about 5 degrees radius), to small (about 2.5 degrees radius). Easy search for a diagonal bar amongst vertical bars became much slower and more error-prone when the size of the visible area was reduced, consistent with the idea that it normally benefits from a large FVF. However, very hard search for a specific configuration of squares hardly suffered at all, even when the visible area was reduced to 2.5 degrees radius – consistent with the idea that for this type of search the FVF was already small to begin with (see Fig. 1 for examples of the easy, medium and hard task used in Young & Hulleman 2013, together with the estimated FVF).

We stress that the idea that the FVF affects visual search is not new. Engel (1977), as well as Geisler and Chou (1995), already showed strong correlations between FVF,

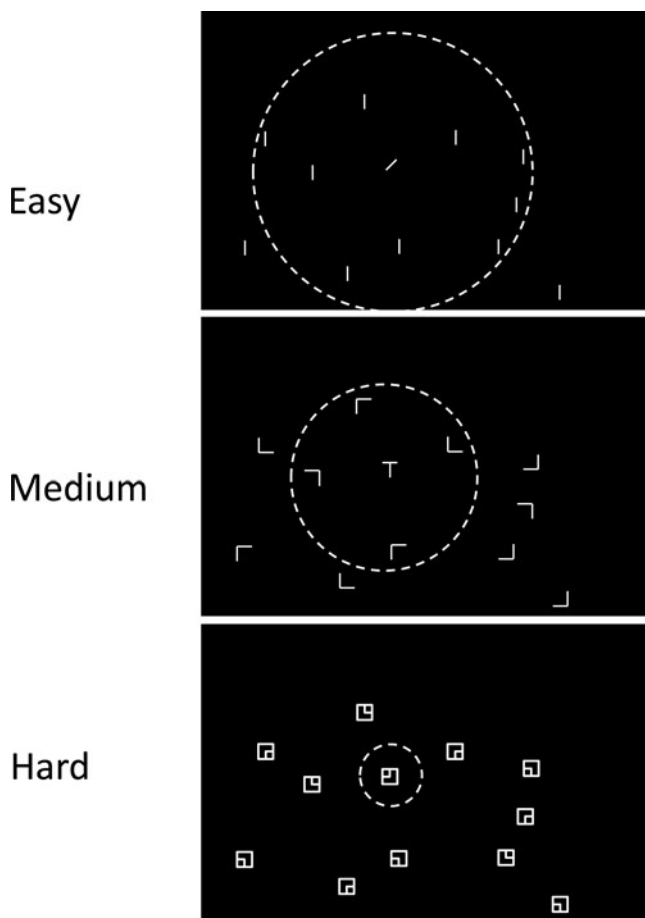


Figure 1. Examples of the tasks used in Young and Hulleman (2013), drawn to scale. Top: easy search for a diagonal amongst vertical; Middle: medium search for a T amongst L; Bottom: hard search for a square with a small square in the left top corner amongst squares with a small square in one of the other corners. The dotted circle represents the estimated FVF for each of the three tasks.

eye movements and overall RTs (although they did not assess RT slopes and variability). Ball et al. (1988) reported effects of age-related changes in FVF on visual search. In his reviews of 1998b and 2003, Wolfe acknowledged the FVF as an alternative way of conceptualizing visual search. Findlay and Gilchrist (1998) also mentioned the FVF as a likely contributing factor to target salience. Nevertheless, somehow the FVF has yet to acquire a firm foothold in mainstream theories of visual search and their computational implementations. We will demonstrate here that the FVF can be considered central to explaining search behaviour, be it eye movements or manual responses. It is not some side effect that at best modulates search but under most circumstances can be safely ignored.

5.2. A simple conceptual framework

As proof of principle we present a simulation of a fixation-based conceptual framework. We deliberately opted for a high-level implementation with only five parameters (four of them fixed), to allow an emphasis on the crucial role that the size of the FVF plays in explaining search slopes and the distributions of both RTs and number of fixations. Thus, the implementation is formulated at a computational level (what are the outcomes) rather than at an algorithmic level (what are the mechanisms; cf. Marr 1982). Specific mechanisms that do not rely on individual items and non-crucial parameters (such as guessing rates and reaction time constants) will need filling in by more detailed algorithms. In fact, some of the details have already been specified in more mechanistic models. What we aim to show here is that connecting the dots leads to a promising overarching theoretical framework.

1. *A functional visual field.* The main assumption of the framework is an FVF. Its size varies with target discriminability. The more difficult the distinction, the less far into the periphery it can be made (Smith & Egeth 1966). For example, the FVF for diagonal amongst vertical is larger than for T amongst L (e.g., Rosenholtz et al. 2012a). As a consequence, fewer fixations will be needed to cover the search display. In our current simulation we have adopted the number of items comprehended at once (c.f. the clumps in Pashler 1987; and the variable number model in Zelinsky & Sheinberg 1995) as a proxy FVF size, although it is properly expressed in terms of visual angle (see Young & Hulleman 2013 for estimates). One may find it odd that we propose an alternative to item-based accounts that is in itself based on items, rather than on a spatial field of view. However, as Figure 1 illustrates, for displays that consist – after all – of items, a spatially limited array of a particular size directly corresponds to a particular number of items. Thus, while using the number of items is a rather crude approximation, it suffices for our current purpose, the simulation of FVFs of different sizes. In our simulation, we assume that the FVF always contains at least one item (the minimum). The maximum it can contain depends on target discriminability. For very easy (“pop out”) search, we assume a maximum of 30 for our displays. For search of intermediate difficulty, the maximum is 7, and for very hard search it is 1. These maximum values were chosen to fit the target-present slopes of the search tasks we simulate. Given that search displays are rarely completely homogeneous, and the FVF certainly is not, the actual number of

items processed within a given fixation fluctuated randomly between the minimum and the maximum, sampled from a uniform distribution. This fluctuation also provides an abstract representation of some of the spatial aspects of the FVF, for instance items that fall in both the previous and current FVF.

2. *Parallel processing of items within the FVF.* Unlike item-based approaches, we assume no additional covert attentional item-to-item shifts within the FVF. Items are assumed to be processed collectively, for example on the basis of pooled statistics (Rosenholtz et al. 2012a) although any other mechanism that selects items as a group rather than individually would be compatible with our framework. Yet, information may come at different rates within the FVF because the FVF is not homogeneous (cf. Findlay & Gilchrist 1998). It is the rate of information accrual in the periphery that is key here, as by definition it determines the FVF, and thus the fixation strategy. Observers may also eventually select individual items. They will do so if the task so requires, for example in compound search. For this, a global search for the target-defining feature may be followed by a local search for the response-defining feature. These different features are likely to have different FVFs, thus requiring different fixation precision. How the system switches between features is currently not captured by the framework, but it does provide a fruitful way of thinking about this problem: as a transition from large to small FVFs.

3. *Fixations of constant duration.* When the FVF does not encompass the entire search display, multiple fixations are required. Fixations are of constant duration, 250 ms. This estimate is based on work reporting only a limited relation between fixation duration and target discriminability in typical search displays¹ (Findlay 1997; Gilchrist & Harvey

2000; Hooge & Erkelens 1996; Over et al. 2007). Fixation duration does not vary with target discriminability. Rather, both the number and distribution of fixations vary with the changing size of the FVF.

4. *Limited avoidance of previously fixated areas of the display.* Avoidance of previously fixated locations improves the efficiency of search (Klein 1988). But visual search has only limited memory for previously fixated locations (e.g., Gilchrist & Harvey 2000; McCarley et al. 2003). Young and Hulleman (2013) also reported revisits to items, even for small display sizes. For the current simulations, we held the number of previously fixated locations that are avoided constant at four (see McCarley et al. 2003). So, given enough subsequent fixations, locations will become available for re-fixation. Because the FVF may contain multiple items, many more than four items might be inhibited during search, because we assume the fixation location, not individual items, to be inhibited.

5. *A stopping rule.* Search is seldom completely exhaustive (Chun & Wolfe 1996). From their eye movement data, Young and Hulleman (2013) estimated that irrespective of display size around 15% of the search items were never visited. For the current simulations the Quit Threshold – the proportion of items to be inspected before search was terminated with a target-absent response – was therefore fixed at 85%. Again, more detailed models have to specify actual stopping mechanisms. Our simulation merely assumes that it is possible to keep track of the proportion of items inspected.

Figure 2 shows a flow diagram of the conceptual framework with its five parameters: Minimum number of items in FVF, Maximum number of items in FVF, Fixation duration, Quit Threshold, and Number of fixation locations

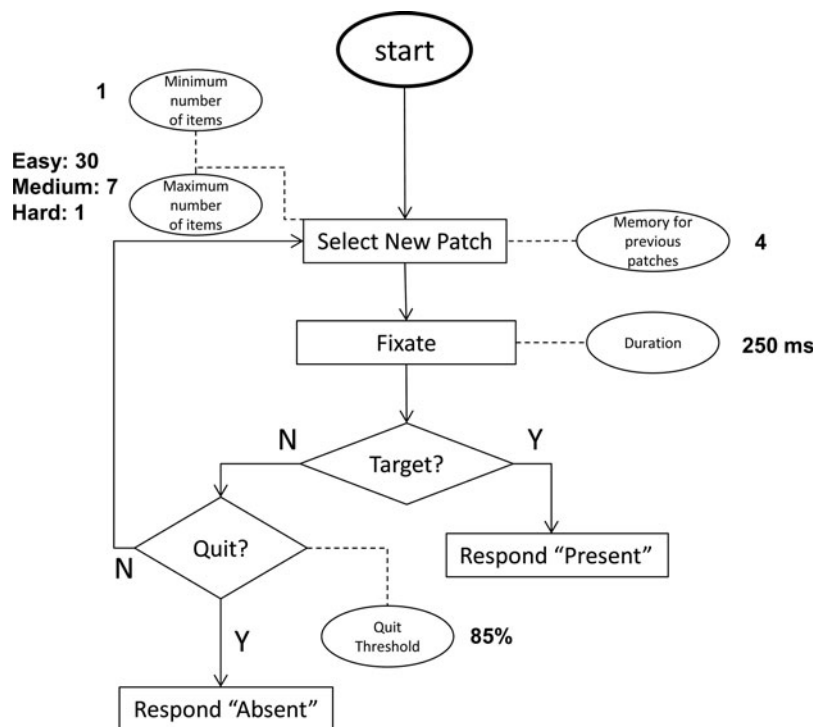


Figure 2. Flow diagram of the conceptual framework. Please see the text for an explanation. For the simulations reported in this paper, the values for the parameters (ellipses in the flow diagram) are printed in bold: Fixation Duration: 250 ms; Quit Threshold 85%; Memory for previously fixated areas: 4. Minimum number of items processed per fixation: 1. For the simulations of easy, medium and hard search, the maximum number of items processed per fixation equalled 30, 7 and 1 respectively.

avoided. The parameters are represented by ellipses, and dashed lines connect them to the parts of the search process that they control. As input, the framework takes the values for the five parameters, plus the number of simulations required and the to-be-simulated display size. Moreover, of these five parameters, *only the maximum number of items in the FVF was variable*. The other four parameters were held constant across all simulations. As output, the simulation gives mean and standard deviations for RTs and number of fixations, proportion error on target-present and target-absent trials and frequency counts for the RTs. Each time a new patch of the display is fixated, the fixation duration is added to the total RT. Thus, the RT for a trial is simply the number of fixations multiplied by their (constant) duration. Although this yields individual trial RTs that are multiples of 250 ms, this is sufficient for our present purposes.

A simulated trial starts by selecting a patch of the display. The current implementation is purely conceptual and does not take images or even lists of features as its input. Instead, search displays are represented as an array of items (effectively, a string of 0's). One of these items is randomly chosen to be the target (and turned into a 1). Again, we point out that for displays that consist of items, this sufficiently simulates a spatial FVF (see Fig. 1). Fixations are implemented as selections from this array, where the number of items selected per fixation depends on the size of the FVF (this number randomly falls between the minimum and maximum for the FVF). If a selection contains the target, search terminates with a target-present response. If the target has not been found, another patch of the display is selected, unless the stopping criterion is met and 85% of the items in the search display have been processed. In this case, the trial terminates with a target-absent response.

We ran Monte Carlo simulations with 10,000 repetitions for each combination of task difficulty, target presence, and display size.

5.3. Simulating the main findings in visual search

The visual search literature is extremely rich, reporting a range of findings too wide to treat comprehensively here. We focus on what we see as the central variables, namely manual RTs and their corresponding search slopes, errors, and fixations, for both present and absent trials. As argued by others, (e.g., Moran et al. 2013; Wolfe 2007), this should be done not only in terms of averages, but also in terms of distributions, because different ranges of behaviour can lead to the same average and thus distributions provide additional information on the underlying process. Of course, any such selection of variables carries a degree of subjectivity, and we refer to Wolfe (2007), for a list of eight core wishes, some of which return here and some others that we regard secondary to the present purpose. We stress again that what we present here is a proof of concept, establishing the viability of our framework, rather than a formal fit of a specific model.

5.3.1. Slopes of average RTs and number of fixations. Figure 1 shows the stimuli used by Young & Hulleman (2013) that yielded data (Figs. 3–6) that in our view are representative for a range of classic visual search

tasks, from very easy, via medium difficulty to outright hard. Exactly by choosing standard laboratory search displays with individual items (rather than e.g., real world scenes), we can demonstrate that our approach is a viable replacement for item-based approaches. We point to Zelinsky and Sheinberg (1995) for an earlier conceptual expression of this idea (though without simulations). Figure 3 shows RT data, Figure 4 shows the SDs for the RTs, Figure 5 shows the number of fixations, and Figure 6 shows the SDs for the number of fixations. Please note the similarity between the patterns for RTs and numbers of fixations. Alongside the experimental data in Figures 3–6, the simulated data are shown. The simulated patterns for RTs and fixations are largely equivalent, because the framework simply assumes that fixations drive the RTs, and fixation duration is held constant. The small differences between RTs and fixation numbers stem from the fact that only correct trials were included for RTs, whereas we included all trials for the fixations (following Young & Hulleman 2013). All in all, with one free parameter, the simulation qualitatively captures the data pattern for both RTs and fixations. For the RTs, it yields flat search slopes in easy search, and intermediate to outright steep search slopes in medium and hard search. Moreover, for both medium and hard search, the slopes are considerably steeper for the target-absent than for the target-present trials. For the fixations, our simulation replicates the finding that target-absent trials in hard search are the only condition where the number of fixations exceeds the number of items (Fig. 5).

The simplicity of our current stopping rule leads to an overestimation of the target-absent slopes in hard search (Fig. 3). Because there is only one item in the FVF and a location is allowed to be fixated any number of times as long as it is not among the last four fixated locations, the simulation has problems reaching 85% by finding the items it has not yet visited, especially for the largest display size. This also becomes clear from the fixations: the number of fixations in the largest display size is overestimated, too (Fig. 5). Clearly, a more sophisticated stopping rule is necessary.

5.3.2 Errors. The simulation yields fairly good estimates for the error rates across the search difficulties. It also captures the increase in error rates for very difficult search (Figure 3).² Because the simulation does not contain a guessing parameter, it always terminates with a target-absent response when it has not found the target. Because this is the correct answer for a target-absent display, the simulation necessarily predicts perfect performance on target-absent trials.

5.3.3 Variability and distributions. The variability of RTs has been problematic for serial item-based accounts, which predict that RTs in target-absent trials will be less variable than RTs in target-present trials. Target-absent decisions can only be made after the last item has been inspected, but a target-present response can be given as soon as the target has been found. Target-absent responses will therefore cluster around the time point of the inspection of the last item. But target-present responses will have more variable RTs, because the target might be the first item selected, the last item selected or any item in

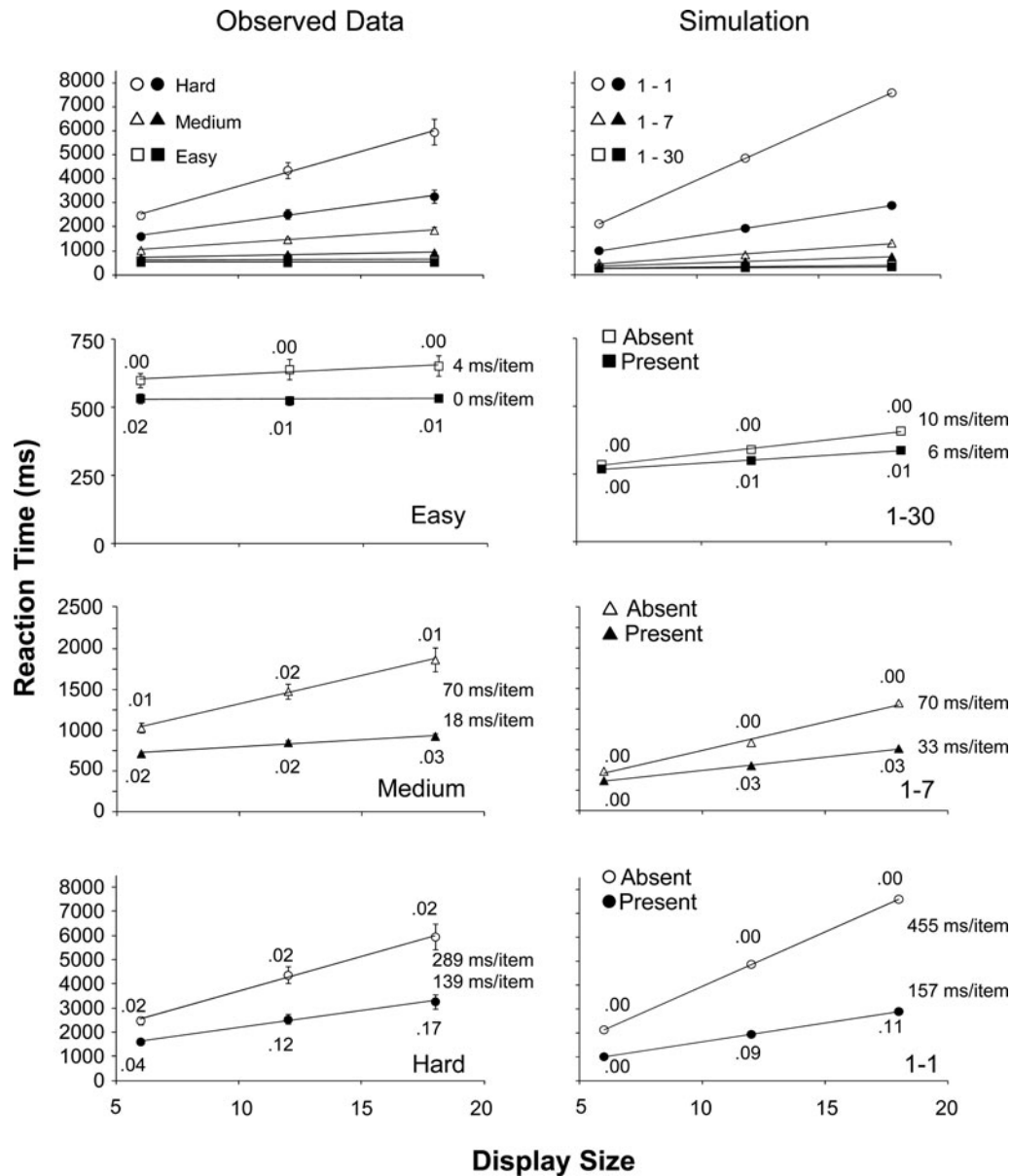


Figure 3. Mean RTs as function of display size. The error proportions are printed next to the symbols and the search slopes can be found on the right of the graphs. Left: Results from Young and Hulleman (2013). The error bars indicate between-participant SEM. In the easy search task, participants searched for a / among |. In the medium difficulty search task they searched for a T amongst L. In the hard search task, the target was a square with a smaller square in the left top corner amongst squares with a smaller square in one of the other corners. Right: Results from the simulation. In the easy search task, 1–30 items were processed per fixation. In the medium task, 1–7 items were processed. In the hard task 1 item was processed per fixation. Top row: overview of the RTs. Second row: easy search. Third row: medium search. Bottom row: hard search. Open symbols: target-absent; closed symbols: target-present.

between. Yet, target-absent trials are typically more variable than target-present trials (e.g., Wolfe 1994; 2007). Guided Search 4.0, has solved this problem, and reproduces the correct variability pattern. Because item-based accounts like Guided Search already consider search for T among L purely item-by-item, they predict that the medium and hard task will both have more variability in target-absent RTs. However, Figure 4 shows that this only holds for medium search. For hard search, RT variability is largest in target-present trials. This pattern is repeated for the number of fixations (Figure 6). Note that this qualitative similarity in the variability of RTs and number of fixations is in keeping with our framework.

As becomes clear from Figures 4 and 6, the simulations capture almost all important aspects of this variability pattern. For hard search, target-present trials show more variability in RTs and number of fixations than target-absent trials, and this difference increases with set size. For medium search, especially at the largest display size, the reverse is found. Here, target-absent trials are more variable for both RTs and number of fixations. For easy search, larger variability for target-absent trials is found for all display sizes.

It is striking that our naïve simulation replicates the reversal from larger variability for target-absent trials in easy and medium search to larger variability for target-present

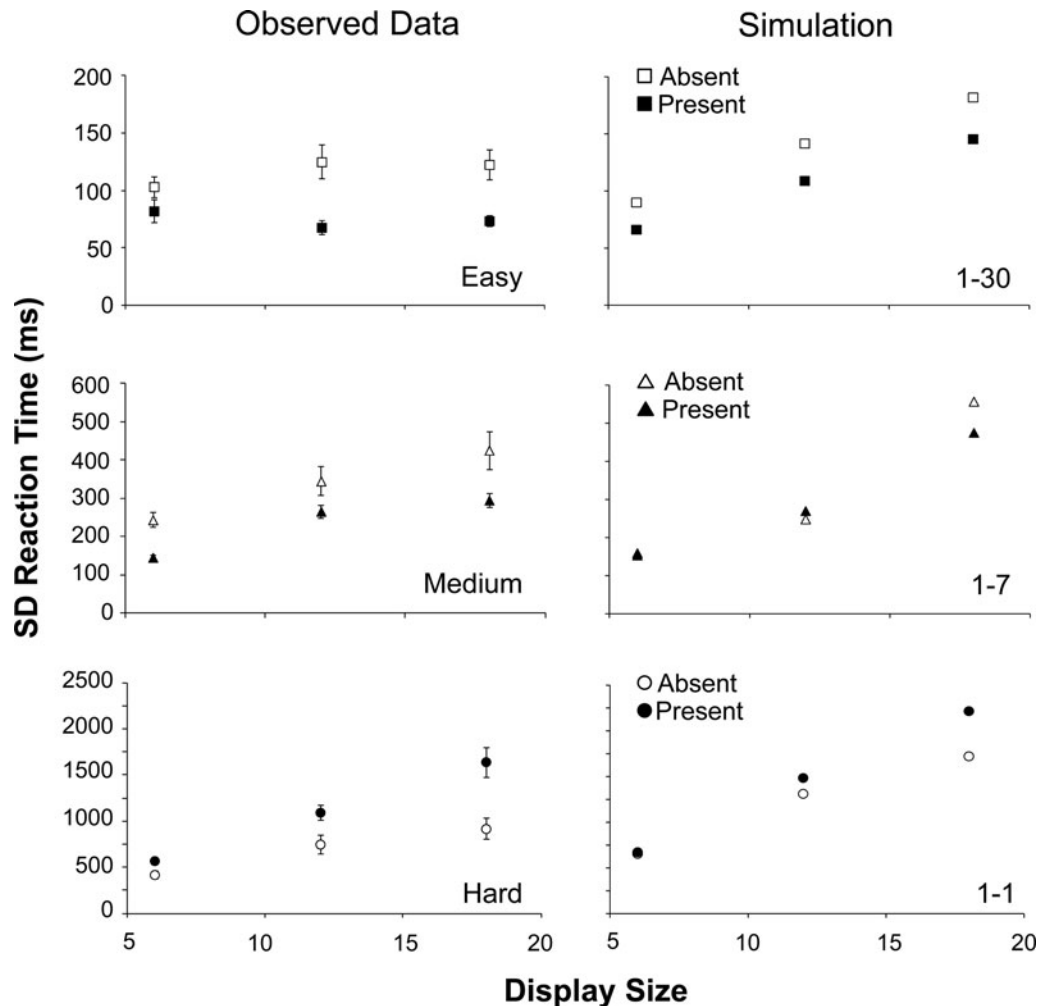


Figure 4. SDs of the reaction times as a function of display size. Left: Results from Young and Hulleman (2013). Right: Results from the simulation. Top row: easy search. Middle row: medium search. Bottom row: hard search. Open symbols: target-absent; closed symbols: target-present.

trials in hard search. This suggests that it succeeds in capturing a crucial aspect of the search process: namely that the size of the FVF increases as search becomes easier.

When the FVF contains multiple items, the difference in RT-variability between target-absent and target-present trials becomes smaller, because the range of time points at which the target is found is reduced substantially. Rather than increasing with the number of items, variability on target-present trials will only increase with the number of fixations. Moreover, the variability of target-absent trials changes less, because the RTs will remain clustered around the time of the fixation that inspects the last items. Consequently, FVFs large enough to contain multiple items selectively decrease the variability in target-present trials.

Furthermore, limited memory for previously fixated locations increases the variability of target-absent trials more. Re-fixations are a source of RT variability, because any number can occur during a search trial. However, the fewer fixations are made to begin with, the fewer re-fixations there will be and the less variability they will add to RTs and number of fixations. Because more time is spent in the search display when there is no target, target-absent trials are more prone to this effect. This combination of FVFs containing multiple items and limited

memory enables our simulation to overcome the inherent tendency of target-present trials to be more variable (because it remains the case that the target can be found during any fixation) whenever search is not too hard.

When search is hard, both these factors lose their influence. First, the FVF contains only a single item. This substantially increases the range of time points at which the target can be found, thereby increasing the variability in RTs for target-present trials. Second, even on target-present trials so many fixations are made that the limited memory for previously visited locations no longer prevents re-fixations.

We also looked at the specific shape of the RT distributions. Figure 7 shows experimentally observed RT distributions for an easy, a medium and a hard task, together with the RT distributions based on our simulations. (We used data from Wolfe et al. 2010a for the easy and medium task rather than the Young and Hulleman 2013 data because the former is based on many more trials per participant.) As Figure 7 shows, the patterns for both sets of data are essentially identical. Across the search difficulties, our simulation captures the distributions for target-present trials fairly well. For easy search, it replicates the narrow single-peaked distribution from Wolfe et al. (2010a),

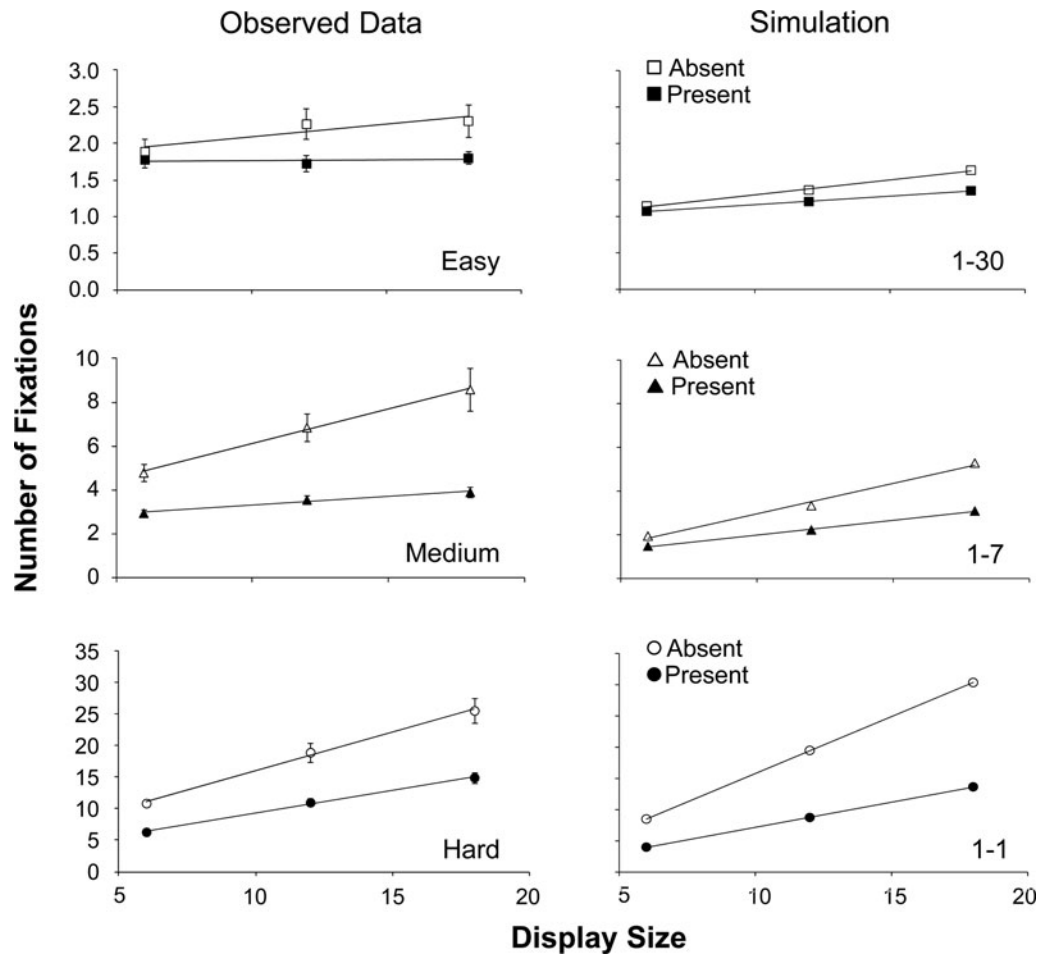


Figure 5. Number of fixations as a function of display size. Left: Results from Young and Hulleman (2013). Right: Results from the simulation. Top row: easy search. Middle row: medium search. Bottom row: hard search. Open symbols: target-absent; closed symbols: target-present. For the results from Young and Hulleman (2013), the error bars indicate SEM.

although there is some widening for the largest display size. For medium search, the simulation replicates both the narrowness of the RT-distribution for the smallest display size and its widening for larger display sizes. For hard search, the simulation yields the relatively narrow distribution of the smallest display size and the very wide distributions for the larger display sizes.

For target-absent trials, although the fits are somewhat less neat, the framework again captures important qualitative aspects. For easy search, the simulation produces the narrow, peaked distribution for the smaller display sizes, although the distribution widens a little for the largest display size. For medium search, the RT-distributions widen with increasing display size, accompanied by a decrease in the mode of the distribution. However, the distributions are not dispersed widely enough around the mode. For hard search, the fit is relatively good: the mode of the distribution for the smallest display size is largest and fairly well-defined, whereas the modes for the larger display sizes are less clearly defined.

The poorer fit of the simulated to the actual target-absent distributions is, again, probably a result of the simplicity of our stopping criterion. Because the target-absent decision is a strategic one, where participants have to take into account a variety of factors, it is unlikely to be

fully captured using a stopping rule as simple as the one used here. We will return to this in the General Discussion.

6. General discussion

There are many different models of visual search, each explaining a fair part of the search process. The most popular ones were designed to account for the mean RTs and error rates observed in search experiments (e.g., FIT, Guided Search, AET). Recent attempts also accurately account for the distribution of RTs (e.g., Wolfe 2007; Moran et al. 2013), while others focused on eye movement patterns in visual search (e.g., Najemnik & Geisler 2008; Pomplun 2007; Zelinsky 2008). So, none of the individual elements of the framework we propose here is new, and much of the hard work has been done by others. However, further progress appears to be stymied. Those models that focus on RTs consider eye movements to be a sideshow. And although we believe that the models that focus on eye movements should be able to account for RT slopes too, so far they appear to have been hesitant to do so (see Zelinsky & Sheinberg 1995, for an exception). We argue that further development of current models is hindered by the implicit but dominant view that the

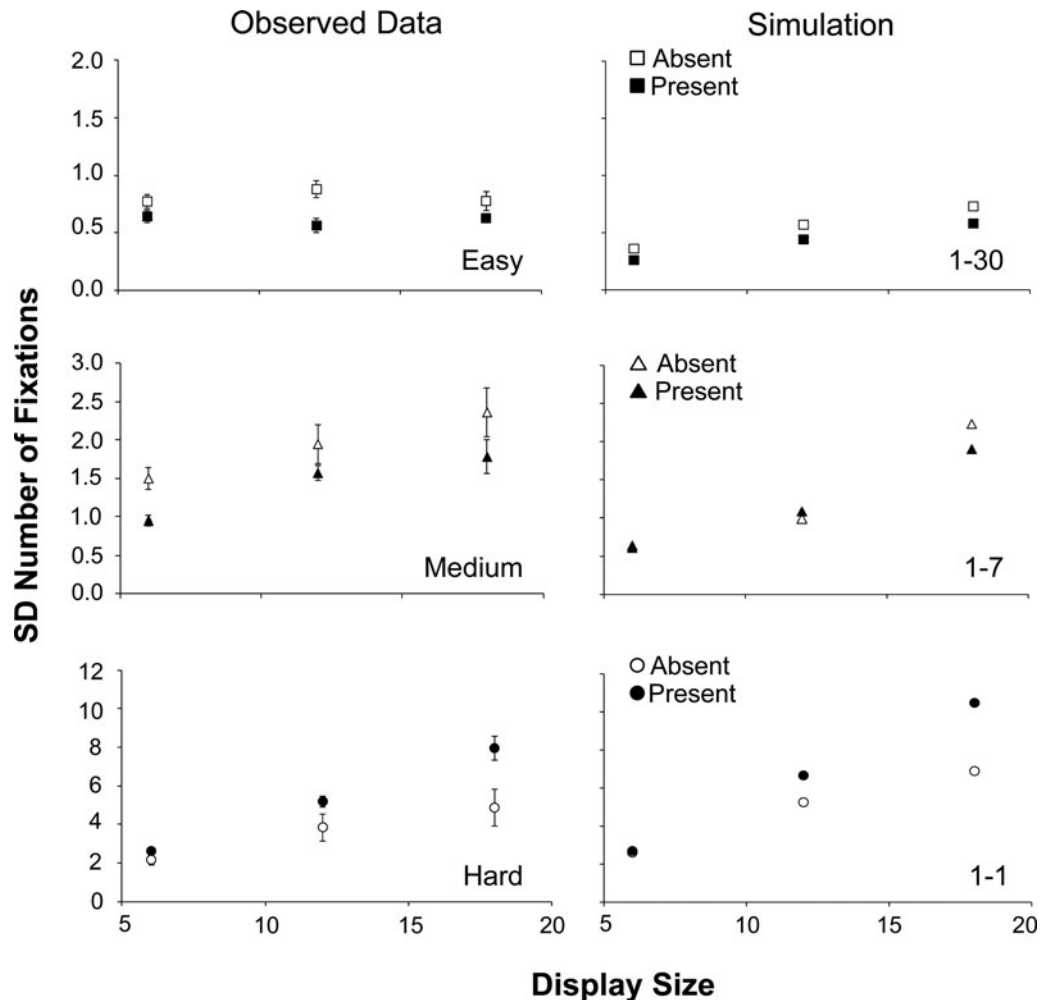


Figure 6. SDs of the number of fixations as a function of display size. Left: Results from Young and Hulleman (2013). Right: Results from the simulation. Top row: easy search. Middle row: medium search. Bottom row: hard search. Open symbols: target-absent; closed symbols: target-present.

individual item is the conceptual unit of visual search, and that therefore (1) the shallow end of the search slope spectrum is the most informative, and (2) eye movements are a nuisance variable to be controlled rather than a crucial theoretical component of the search process. These implicit assumptions have prevented current models from explaining all aspects of search, from eye movements to manual RTs, from errors to distributions, from present to absent responses, and from very hard to very easy search. We hope that our simulation has shown that an approach to visual search that abandons these implicit assumptions has a lot of descriptive power, and holds the promise of uniting the wide range of findings into a single framework. With only five parameters, the framework provides integrated and qualitatively accurate predictions of the means and variability of RTs and number of fixations, as well as error rates for a variety of search tasks. The framework thus appears to have achieved the goal of uniting manual and oculomotor behaviour.

In particular, our simulations suggest that rather than being a reflection of the selection rate of individual items, search slopes can be the consequence of the interaction between a constant fixation duration and the size of the FVF. The simulation shows that with appropriately

chosen FVF sizes, the framework covers the entire range of observed search slopes, from 0 ms/item to 150 ms/item for target-present, without having to change any other fundamental assumptions. Despite earlier debates (Eckstein 2011; Findlay & Gilchrist 1998; Zelinsky & Sheinberg 1995), the emphasis on items as the conceptual unit in visual search seems to have led to an enduring misinterpretation of search slopes. We believe our simulation now provides a compelling argument for abandoning the stance that search primarily revolves around individual items.

6.1. RTs and variability

According to Wolfe et al. (2010a), successful models of visual search should describe both mean RTs and their variability. To capture the fact that target-absent trials are more variable than target-present trials in a search of moderate difficulty (for 2 amongst 5), Guided Search 4.0 had to adopt a new item selection mechanism compared to its predecessors. However, because it is item-based, the model also predicts that larger variability in target-absent trials should be found across the entire range of search difficulties. But when search becomes very hard there is a crossover, and target-present trials become more variable.³ It

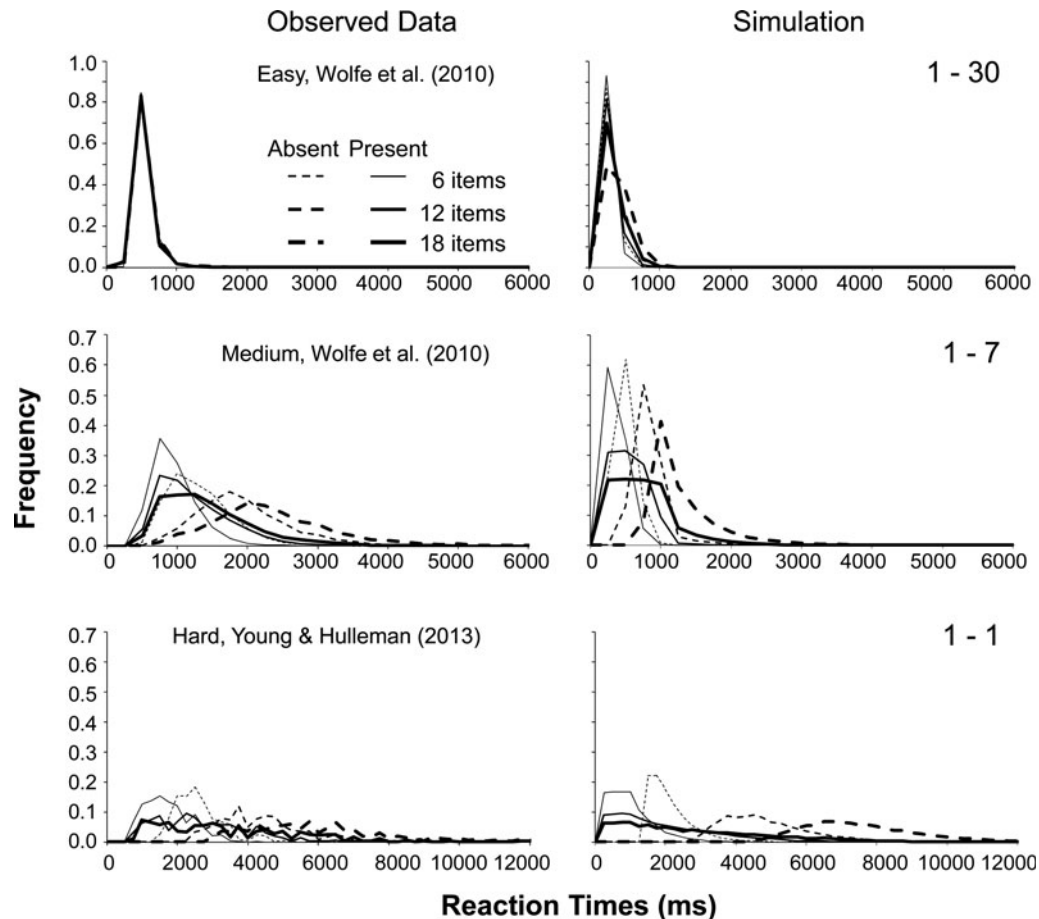


Figure 7. Left: Experimentally observed RT-distributions. Right: Simulated RT-distributions. Top: easy search (red vertical amongst green verticals, search slopes 1.0 and 0.7 ms/item for target-present and target-absent, respectively) reproduced from the Wolfe et al. (2010a) data set. Simulated FVF: 1–30 items. Middle: medium difficulty search (spatial configuration search -2 amongst 5's-, search slopes 43 and 95 ms/item for target-present and target-absent, respectively) reproduced from the Wolfe et al. (2010a) data set. Simulated FVF: 1–7 items. Bottom: Hard search task (square with smaller square in left top corner amongst squares with smaller square in other corner, search slopes 139 and 289 ms/item for target-present and target-absent, respectively) based on the data of Young and Hulleman (2013). Simulated FVF: 1 item. Solid lines: target-present trials. Dashed lines target-absent trials. Bin size: 250 ms.

is especially this crossover in variability that is not trivial to capture in item-based models of visual search. Recently, Moran et al. (2013) have presented an extension of Guided Search that is also capable of reproducing the RT-variability patterns that are observed in medium difficulty search and easy search (albeit with separate sets of 8 fitting parameters each). It would be interesting to see whether the same model can cope with the inversion of the RT-distributions in very hard search. Here our framework appears to make unique predictions.

We hold that the difficulties or even the failure of these models to capture the distributional aspects of RTs in visual search is a direct consequence of the implicit assumption that the individual item is its conceptual unit. Our simulation shows that this may be the case for a very hard search, but for most searches, including that for T amongst Ls, the data pattern is best captured if one assumes that the FVF covers multiple items. This assumption allows the other factors that influence termination of target-absent trials to increase RT-variability to such an extent that it becomes larger than in target-present trials. By themselves these factors are not enough, as is demonstrated by the larger variability in target-present trials when the FVF

only covers a single item. Thus, the adoption of the individual fixation as the conceptual unit offers a promising way to capture RT-distributions (and distributions of number of fixations for that matter) across the entire range of search difficulties.

Even if item-based models were to successfully fit the RT distributions and the error rates across a wide range of search difficulties, we would still argue that fixations should be preferred over an item-based selection rate because this choice increases explanatory power while maintaining relative simplicity. There is a direct link between number of fixations and RTs in our framework. The rate of selecting new parts of the search display is fixed at 250 ms, and has a clear basis: it is how long the eyes remain steady. This sets it apart from models of rapid serial covert shifts of attention like Guided Search and AET. These would need additional assumptions to incorporate eye movements. The need for a direct link becomes clear when RTs are plotted as a function of number of fixations (Figure 8): There is an almost perfect linear relationship, irrespective of set size, presence of the target, or difficulty of the search task (see also Zelinsky & Sheinberg 1995). For our framework, this is only natural,

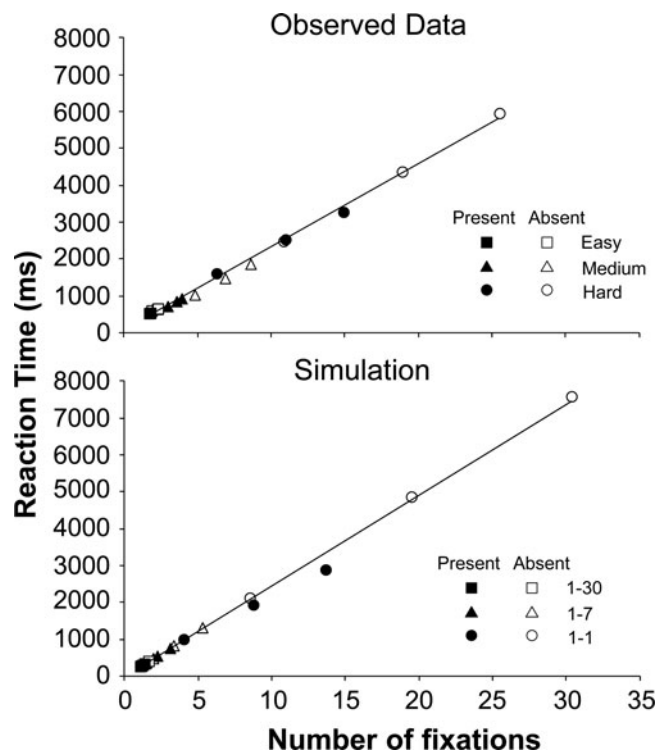


Figure 8. RTs as a function of fixation. Top: data from Young and Hulleman (2013). Bottom: simulated data. Open symbols: target-absent. Closed symbols: target-present. Squares: Easy search / 1–30; Circles: Medium search / 1–7; Triangles: Hard search / 1–1.

even trivial, because RTs are directly based on the number of fixations. For item-based accounts though, this linearity is not so trivial, because they explain differences in search slopes through differences in covert or central selection rates that are in principle independent of eye movements. There is therefore no *a priori* reason for the RTs from a range of task difficulties to line up so neatly.

6.2. The benefits of the Functional Viewing Field

Of course, within our framework the assumption of an item-based selection rate is replaced with another assumption, namely the FVF. One could argue that we are simply robbing Peter to pay Paul. However, we believe the concept of the FVF to be more elegant for several reasons.

First – and this is our main point – the FVF allows for an integrated explanation: Search RTs are determined by the number of fixations. The number of fixations is determined by the physiological limitations of peripheral vision, in terms of both retinal resolution and neural competition in early sensory areas. In other words, rather than being limited by some central processor, search is limited by the eye.

Second, an emphasis on the FVF allows for direct links to other findings in visual perception research, such as the crowding and lateral masking literature. So far, many visual search investigators tend to avoid these phenomena. By accepting a direct link between them (as was already done by Engel 1977; Geisler & Chou 1995; see also Motter & Belky 1998b), we can start to investigate *how* they affect visual search. Taking into account the peripheral

limits of vision might also explain part of the seemingly coarse coding of orientation in visual search (e.g., Wolfe 1994): orientation discrimination thresholds increase with foveal distance even when only a single stimulus is presented (Westheimer 1982). More emphasis on the FVF would likewise encompass data from Duncan and Humphreys (1989). In addition to reducing the grouping which allows distractors to be rejected simultaneously (as suggested by Duncan & Humphreys 1989), increased target-distractor similarity probably renders peripheral detection of the target more difficult, thus reducing the FVF and increasing the need to foveate items more closely. This increases the number of fixations needed to cover the display and therefore the RTs. Similarly, increasing the similarity between distractors increases homogeneity, making peripheral targets more discriminable. This increases the FVF and thereby reduces the need for eye movements and decreases RTs. In this sense, the FVF could be seen to implement Duncan and Humphreys' (1989) original idea of “structural units”: groups or chunks of visual objects defined at different scales that constitute the input to the search process.

Third, differences in FVF size explain the difference in robustness against item motion between medium search and hard search (Hulleman 2010). Larger FVFs offer protection against item motion in at least two ways. First, in larger FVFs, multiple items are processed simultaneously, so any re-inspection of an item that moved from a previously fixated location to the currently fixated one does not incur that much of an RT-cost. Second, because larger FVFs yield fewer fixations, there will be fewer re-inspections to begin with. When the FVF contains only a single item, both types of protection disappear. When an item is re-inspected, there is a substantial RT cost, and because many more fixations are needed when only a single item is processed in the FVF, these re-inspections are more likely to occur.

Fourth, it allows us to move on from using visual search to diagnose which visual properties deserve special feature status (a research line to which we happily contributed ourselves, e.g., Hulleman et al. 2000; Olivers & Van der Helm 1998 and, though dwindling, is still ongoing see e.g., Li et al. 2014). Assigning feature status implies a binary classification according to which the visual property either is or is not available for guiding attention (as implemented in separate feature maps). Guidance is then expressed through shallow search slopes. We do not argue against the existence of basic features, and the original guiding features such as colour and orientation have clear connections to cortical architecture. However, the criterion of flat search slopes has yielded a wide variety of candidates, some quite unlikely (e.g., pictorial depth cues, shading, see Wolfe & Horowitz 2004 for a complete overview). Wolfe and Horowitz (2004) therefore argued that flat search slopes should not solely determine feature status. After applying further criteria, Wolfe and Horowitz (2004) accorded undoubted feature status only to colour, motion, orientation and size.

FVFs allow for a wider range of visual characteristics to come into play. By accepting the FVF as the major delimiter, the question as to what is a feature can be replaced by what is detectable from the corner of the eye. This is likely to correlate with the existing feature rankings, but allows for more flexibility, as detectability will improve with any sufficiently large difference signal relative to the surround, whether feature-based, conjunction-based, or based on

more widely sampled statistics. Thus, where Guided Search explicitly dissociates guidance from discriminability (Wolfe 2007), the FVF account predicts that any discriminable property can be used for guidance, even Ts amongst Ls. Flat search slopes no longer confer special status, but simply indicate that a particular discrimination is easy (or target-distractor similarity is low, cf. Duncan & Humphreys 1989). Moreover, factors such as relative size (scaling) are also naturally allowed into the equation. Instead of the binary distinction between features and non-features, FVFs thus offer a continuous spectrum of search performance.

Finally, we tentatively propose that the concept of an FVF also allows for a more integrated explanation of semantic biases in search through scenes. Research has indicated that participants are able to bring to bear their knowledge of scenes when they have to search for a particular item (e.g., a pan in a kitchen, Henderson & Hollingworth 1999; Wolfe et al. 2011a). They mainly fixate likely parts of the search display (i.e., horizontal surfaces) and avoid unlikely ones (e.g., the walls and the kitchen window). Within purely item-based approaches to visual search, where individual items are selected at a rate of 25–50 ms/item, this semantic guidance is difficult to explain. Establishing that the scene is a kitchen to begin with seems to require the painstaking collection of multiple items. Yet, research shows that very brief exposures of 50 ms are enough for scene-categorization (Greene & Oliva 2009). To account for this fast scene categorization an entirely separate pathway for the parallel processing of scene information has been invoked to bring search of scenes within the purview of models of “classic” search like Guided Search (e.g., Wolfe et al. 2011b). This pathway rapidly determines the scene type, and then passes on this information to bias the search process in the other pathway via semantic and episodic guidance towards likely spatial locations. Yet the underlying search mechanism has not changed. Individual items are still selected as candidate targets. A separate parallel scene pathway is unnecessary for a fixation-based account like the one proposed here. The FVF already assumes parallel processing, and allows for the extraction of global information on the basis of image statistics, at the very first fixation. While the underlying computations would remain similar, different types of information will yield different FVFs. Whereas the FVF for an individual object in a scene may be small, the FVF for recognizing the scene as a whole is likely to be much larger. Thus, the same computations across a smaller FVF that allow the decision that a T is present in a display full of Ls may be used across a very large FVF to establish that this is a forest scene rather than a city scene, or that an animal is likely to be present rather than a car (c.f. Thorpe et al. 1996). An interesting question for the future, then, is the precedence of various FVFs for different types of information.

6.3. Can item-based models not be easily adapted?

We believe extending item-based models to include eye movements will be challenging. Note that Guided Search, as the most important and developed proponent of item-based search, goes quite a long way in trying to maintain the individual item as the core of the search process: It postulates a separate global scene processing pathway in order

to preserve the item processing pathway, it assumes a selection bottleneck to connect fast item selection at the front-end to slow item processing at the back-end, and it considers eye movements at best as the reason that there is such a slow item processing bottleneck. Moreover, the separation of item selection from item processing that has been implemented with the car wash mechanism of Guided Search 4.0 (Wolfe 2007) creates problems when eye movements are taken into account. Important differences emerge between the item that enters the processing stream *first* (the processing of which will be nearing completion by the time of the next fixation) and the item that enters *last* (the processing of which will only just have started). For example, an experiment reported by Henderson and Hollingworth (1999) suggests that the representation of an item deteriorates once it is no longer fixated. Participants were less likely to detect a change made to an item during a saccade if the item in question had been fixated immediately before the saccade. If items are selected sequentially, this deterioration of representation would affect the most recently selected item much more than the item selected first. Moreover, single-cell recordings in the lateral intraparietal area of monkeys (Kusunoki & Goldberg 2003) show a reduction in sensitivity of the receptive field even before a saccade is made. This finding suggests that the representation of items might already deteriorate towards the end of the fixation. Again this implies that, if items were selected sequentially, items that enter the processing stream early will have an advantage over those that enter the processing stream late. An approach based on fixations as proposed here allows these selection order problems to be circumvented, because all items, in principle, are selected and processed simultaneously. A further simplification offered is that selection time, processing time and dwell time are all allowed to be identical and equal to the fixation duration. In other words, our framework makes additional assumptions about the role of central selection bottlenecks redundant.

6.4. What about covert search (when the eyes are kept still)?

It might seem that our account is fundamentally flawed simply because it is possible to complete visual search tasks without eye movements, and results seem identical when corrected for peripheral limitations. There are several answers to this objection. First, our account takes *fixations* as units, not eye movements, and every search includes at least one fixation. Our simulation allows for target-present responses during the first fixation. Figure 7 shows that most of the target-present responses in easy search did not involve a second fixation, and even for hard search there are trials with only one fixation. For easy search there are also many target-absent responses that do not involve eye movements. However, it is indeed the case that medium and hard search typically involve multiple fixations.

Second, and more important, crucial to our fixation-based framework are the presumed limitations of the FVF (which under normal circumstances lead to eye movements), not the eye movement per se. Thus, even if no eye movement is made, the non-homogeneity of the visual field in terms of attention and lateral masking is still very likely to influence selection. Even if targets can in principle be

detected from the corner of the eye when eye movements are not allowed, it is still likely that detection takes longer or becomes more erroneous when targets become less discriminable and set sizes increase. In fact, the argument that search can proceed covertly at a level equivalent to overt search when displays are *corrected* for peripheral limitations implies that usually covert search is much harder. Related to this, one reason why participants may choose to make eye movements in the first place is that although they *could* perform a particular task far into the periphery, it might simply be more comfortable or more efficient to make an eye movement that moves the area of scrutiny closer to the fovea, even though this in itself might take some time (Findlay & Gilchrist 1998). Thus, search with and without eye movements may be similar because they are driven by the same FVF.

Some results appear to support this contention. For example, Klein and Farrell (1989) compared search performance with and without eye movements. Although search latencies were nearly identical, analyses of the error patterns showed that, without eye movements, participants encountered difficulties with the larger display size, particularly on target-absent trials, where error rates doubled. Thus, it appears that although the task could be performed in principle, targets did become harder to discriminate when eye movements were not allowed. Zelinsky and Sheinberg (1997) found little difference in error rates between the fixed-eye condition and the free-eye condition for their difficult search tasks. However, this time the fixed-eye condition yielded faster RTs, especially for the largest display size. One reason for the relative disadvantage in the free eye movement condition may lie in saccadic suppression, the loss in visual sensitivity during an eye movement (e.g., Volkman et al. 1978). Another may be the more widely spaced displays used, which contained a relatively large number of items. Although all items may then in principle be visible *initially* from the central fixation point, once observers make an eye movement to an area at one end of the display, they lose acuity (and conspicuity) for the items at the other end, potentially to the point that the target is no longer discriminable, making further eye movements necessary. If so, making one eye movement likely results in making more. Thus, whether free viewing provides a benefit or a cost depends on the design of the display, which determines the FVF (see also Findlay & Gilchrist 1998, for a similar point).

6.5. Does this FVF approach make any predictions at all?

One criticism of our approach might be that it does not make any predictions beyond the simple observations that if search takes long, the FVF must have been small and if search is fast, the FVF must have been large. It does not contain an independent mechanism that predicts the size of the FVF.

Our first response to this is that it is not our aim to present a mechanistic model. Rather, we present a conceptual framework, to open up different ways of thinking about visual search, and the important new questions this raises – together with a conceptual demonstration that it works. Indeed, clearly, what determines the size of the FVF is one of the main questions arising from the framework. But equally clearly, this is not a question that easily

follows from an item-based approach, in which the main questions are about the way individual items are processed.

Second, our approach does identify where to expect the largest differences. For example, a direct prediction of the model is that there should be qualitative differences between hard search on the one hand, and easier search on the other. Some we already identified here: robustness against motion, the influence of peripheral information, the relation between number of fixations and number of items, and the pattern of RT distributions. Others have yet to be explored (e.g., the effects of age-dependent changes in the size of the FVF; Ball et al. 1988; Sekuler et al. 2000).

Third, existing item-based approaches do not escape circularity themselves. For example, whether a particular property is a feature or not is determined by the search slope. In AET, search efficiency is determined by target-distractor and distractor-distractor similarity, which so far have been only been expressed through changes in the search slopes, rather than on the basis of independent data or a particular mechanism (with Treisman, 1991, as an exception). It actually could be argued that an approach based on FVFs holds the best promise for an escape from circularity, because there is a direct correlation between search slopes and the outcome of single fixation detection tasks outside search (Engel 1977; Geisler & Chou 1995).

6.6. Remaining questions and future directions

Simple and naïve as our framework is, it points out some clear areas of further research. As alluded to above, the first knowledge gap that needs filling-in is what determines the size of the FVF during search. Why do some tasks have larger FVFs than others? This is an important question, especially because it seems likely that the factors that determine FVF size will be closely intertwined with the way target-presence is determined within the FVF. In our simulation, the difficulty of the search task was assumed, but preferably the assessment of task difficulty should be based on properties of the search display. In pixel-based approaches, this problem is computationally tractable. For example, in TAM (Zelinsky 2008) task difficulty is determined from the search display by establishing how much more target-like the most promising area of the search display is relative to other areas. If the difference is large, the search task is easy and large eye movements can be made. If it is small, the search task is difficult and smaller eye movements should be made. Presumably, a more homogeneous distractor set will make the target area stand out more, in line with the central role that target-distractor similarity and distractor-distractor similarity play in AET (Duncan & Humphreys 1989). Thus, TAM provides a promising avenue for independently assessing discriminability.

A different approach with similar outcome may be the computation of summary statistics across the items in a fixated patch of the display, as suggested by Rosenholtz et al. (2012a). These summary statistics allow the visual system to establish whether the patch contains a target or not. A salient target against a homogeneous background will create a reliable deviation in the summary statistics, while a weak signal against a noisy background will deliver unreliable statistics. More reliable signals allow for larger patches to be sampled, and thus larger FVFs. As mentioned earlier, some types of search tasks require more than can be delivered by large FVFs alone, as a specific target property is

required (i.e., its exact location or independent response feature). This means that FVF size may be dynamic, changing on-line from large when the target has to be acquired to small when a specific aspect of the target has to be reported. Such changes have yet to be explored.

A related question is how the next fixation location is selected. In our simulation, it was simply assumed that new items would be selected. Determining the next location to be selected may depend on many different factors, but here eye movement models have a clear advantage over RT-models, because they are specifically designed to describe eye movement patterns. In fact, there already are several candidate mechanisms differing in the way they choose the next location. For example, TAM (Zelinsky 2008) computes the correlation between a target-template and each pixel in the search display. This correlation will be highest for the pixel at the centre of the target, but other positions can also have high correlations. The next fixation location is determined by computing the average position of all of the correlations. By increasing the minimum correlation that is required to contribute towards this average, fewer and fewer pixels will contribute and the average position starts to move towards the target. When the average position exceeds a distance threshold from the current fixation location, a saccade is triggered. So, fixation positions are chosen based on the most likely target candidate. However, fixations do not necessarily fall on an item, because the average of all contributing pixels in the display does not necessarily coincide with an item.

In a different approach, the ideal searcher of Najemnik and Geisler (2008) computes the probabilities that the target is located at a number of candidate locations, by correlating the locations with a target template. It determines the optimal next fixation location by trying to maximize the possibility of identifying the target *after* the next fixation, taking into account the current probabilities and acuity limitations. Therefore, fixation locations are chosen based on which location will be most useful in the subsequent decision where the target is, rather than the current most likely target location. A related suggestion comes from Pomplun et al. (2003; Pomplun 2007) who proposed the Area Activation Model (AAM). AAM computes the relative informativeness for each fixation position, by taking a weighted sum of all of the surrounding guiding features. The informativeness depends on search task difficulty, with a larger area of the display contributing to the informativeness measure when the task is easier. The resulting map has several peaks of informativeness, which do not necessarily coincide with individual items but could fall in between. The first saccade will go to the highest peak, the next saccade will then go to the nearest non-visited peak.

Both TAM and AAM make allowances for the difficulty of the search task when choosing the next fixation location. But an interesting aspect of AAM is that it rejects groups of items when the group of items contributing to the informativeness peak does not contain the target. This makes AAM more compatible with our proposed framework than TAM, where always only a single item is matched to the target template and inhibited when it turns out to be a distractor, irrespective of the difficulty of the search task.

The third question is how trials are terminated when no target has been found. All models of visual search, including the framework presented here, seem to be much better at describing target-present trials than target-absent trials.

Because for many critical tasks the consequences of a miss are much more severe than the consequences of a false alarm (X-rays and CT-scans, airport security screening), it is also from an applied point of view vital to understand target-absent decisions better. Target-absent decisions not only influence the RTs for target-absent trials, but also the error rates on target-present trials. Both of these areas are amongst the weaker aspects of the framework we have presented here. They are amongst the weaker aspects of mechanistic models (e.g., Guided Search) too and some leave out target-absent trials altogether (TAM; Zelinsky 2008, but see Zelinsky et al. 2013). Multiple triggers for the target-absent decision have been proposed (number of items inspected, time spent searching the display, success of previous target-absent response, frequency of target-presence; see also Chun & Wolfe 1996), but they all seem to be weighed at different rates at different times, without a clear ranking of their importance. Any simple model of target-absent decisions (and, therefore, any simple model of visual search) seems doomed to fail in its attempt to capture the essence of target-absent decisions, especially when the entire spectrum of search difficulty has to be taken into account. This is demonstrated by our simulations. Our simple stopping-criterion terminated medium searches too early, but at the same time let them continue too long in both easy and difficult search. At the very least, this suggests that participants do weigh the difficulty of the search task in their target-absent decision. In that sense, future mechanistic models can be improved by letting task difficulty not only shape the search process by determining the size of the FVF, but also by changing the criteria for terminating a search when the target has not been found.

Finally, some of the most exciting areas in visual search are tasks in medical imaging (does this mammogram contain a suspicious lesion) and in airport security (does this bag contain a threat). Although these areas have seen considerable interest from fundamental cognitive psychology (Donnelly et al. 2006; Drew et al. 2013a; 2013b; 2013c; Evans et al. 2013a; Godwin et al. 2010; Menneer et al. 2007; Wolfe et al. 2005; 2013), they have been underserved by item-based models (or any form of overarching theory of search, for that matter), not least because it is difficult to determine how many items the kind of images typically used actually contain. Moreover, the kind of target that needs to be found (“threat,” “lesion”) also sits uncomfortably with an item-based approach, because they can take on a multitude of fuzzy forms and are therefore difficult to capture in a target-template. Models are necessarily tested by deriving predictions from them. Item-based models will make item-based predictions and it is therefore only natural that many lab-based experiments use displays with clearly defined items. Unfortunately, this has opened up a gap with real-world tasks that is only now beginning to be bridged. We believe that by emphasising the role and importance of fixations, this bridging process will be sped up, because it focuses on a factor that lab-tasks and real-world tasks have in common. We hope that our proposed framework can be a starting point, which allows the exploration of the many factors that influence real-world tasks (experience, time pressure, age, target-prevalence, training, unknown number of targets, complex backgrounds) while at the same time providing a foundation for more fundamental research into the processes underlying visual search, bringing real-world tasks and lab tasks closer together.

7. Conclusion

Our simulation has demonstrated that a fixation-based framework shows considerable promise as an integrated account of visual search, and allows for the abandonment of some of the implicit assumptions that have dominated the field for decades. It reveals how an acknowledgement of the functional visual field and the adoption of the number of fixations (rather than the number of items) as the driver of RTs yield a surprisingly adequate description of various aspects of visual search behaviour. Although the conceptual nature of the framework is an obvious weakness, it is also a core strength: Exactly because the framework does not specify the details of the mechanisms involved in visual search, it allows a clearer view of the explanatory power of the underlying principles.

ACKNOWLEDGMENTS

The title of this paper was inspired by Ralph Haber's (1983) "The impending demise of the icon: The role of iconic processes in information processing theories of perception" (in *Behavioral and Brain Sciences* 6:1–11). This work was funded by Leverhulme Trust grant F/00 181/T awarded to JH, and by ERC Consolidator Grant 615423 awarded to CNLO. We thank Jan Theeuwes, Martijn Meeter, Andrew Hollingworth, Ruth Rosenholtz, and Derrick Watson for reading an early version of the manuscript, as well as four anonymous reviewers. We would also like to thank Jeremy Wolfe for making his data available.

NOTES

1. When search items are highly surprising or very difficult discriminations have to be made *within* an item, fixation durations might increase.

2. For the difficult condition, search is terminated closer to the quit proportion. Because items are inspected one by one, the number of inspected items will exceed the limit imposed by the Quit Threshold by maximally 1. For medium and easy search, 1 to 7 and 1 to 30 new items are inspected per additional fixation. Therefore the limit based on the quit proportion can be exceeded quite substantially. Consequently, more items are inspected on average and fewer errors will be made. The larger FVF is also the reason that error proportions for easy search are even lower than for medium search.

3. Admittedly, we draw mainly on our own work for this observation. This is due to the rarity of other studies that look at the RT variability in hard search. This is probably exactly because standard item-based theories hold that hard search does not add to the observations in medium search, and is furthermore complicated by eye movements.

Abstract: Rejecting information-processing-based theory permits the merging of a top-down analysis of visual search tasks with a bottom-up analysis of brain structure and function. This reveals the true nature of the functional visual field and its precise role in the conduct of visual search tasks. The merits of such analyses over the traditional methods of the authors are described.

Aims. The authors have served the science well by introducing the important concept of the functional visual field (FVF), but they fail to exploit it fully because of their adherence to conventional ways of working, namely embracing an information processing view of cognition, using data from invented laboratory tasks, and basing explanations of performance on post hoc conjecture rather than strong theory.

I here describe research based on new ways of working that develops a strong theory of the FVF and its role in real-world vision. In doing so it provides answers to three questions posed by the authors: How do covert shifts of attention and eye movements relate to the FVF; what determines the size of the FVF; and how are fixation points selected?

To date the work has been presented only piecemeal for rather niche audiences; more comprehensive mainstream treatments are in preparation.

Methods. Theoretical and methodological challenges encountered in early neuropsychological work (Campion 1987; Campion et al. 1983), and in later applied work (Campion 1989), led me to develop a theoretical paradigm that, unlike information processing, was able to encapsulate what the world actually *looks like* to perceivers, and how this depends on what they are trying to do in it.

This laid the foundations of a general theory of visual perception based on a reworking and integrating of ideas from classical approaches such as *Gestaltism* (Koffka 1935), *Ecological Optics* (Gibson 1979), and *Constructionism* (Neisser 1966) – ideas which information processing had destroyed through assimilation (see Campion 2014; Palmer 1999). These reworked ideas were blended with modern conceptions of brain structure and function (Campion 2009; Campion 2011).

Four core ideas underpin the theory:

1. Perception is not the *processing of information* but the *instantiation of knowledge* – the controlled, energy-consuming fusion of a sensory database with learned knowledge.

2. Knowledge is of different types and is constructed and recruited by situation-sensitive learned plans to guide the conduct of tasks.

3. Cognitive and neural processes are not decoupled as information processing doctrine maintains (e.g. Marr 1982), but are different *levels of description* in the reductionist sense.

4. Neural tissue does not perform computations, but has specialised soft-wired circuits that are established and consolidated through experience and are switchable according to task demands.

It follows that the brain can be properly understood only by identifying the various levels of description, and establishing the mappings between them. The case is made here, not by a priori argument, but by demonstrating that it works.

The top level identified is the *task*. A cognitive task analysis (CTA) of the line cancellation task used to diagnose visual hemispatial neglect (Albert 1973) was blended with an analysis of literature on brain structure and function. The use of CTA is an important innovation here, but is standard fare in cognitive ergonomics where this work originates (Crandall et al. 2006).

Results. Research suggests that the *configural* aspects of the world are handled by two knowledge systems that are distinct but that work together – a *locating* system subserved by the right parietal region (RPR) of the brain (see Karnath et al. 2002), and a *manipulating* system subserved by the left parietal region (LPR) – see Sirigu et al. 1999. The former is the focus here.

Open Peer Commentary

Analysing real-world visual search tasks helps explain what the functional visual field is, and what its neural mechanisms are

doi:10.1017/S0140525X16000030, e133

John Campion

30A London Road, Liphook, Hants GU30 7AN, United Kingdom.
tj.campion@btinternet.com

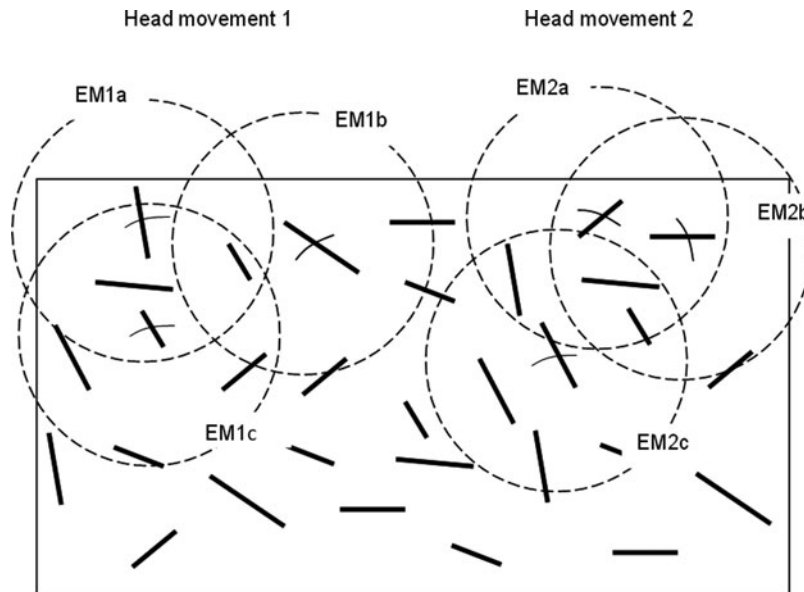


Figure 1 (Campion). Analysis of the line cancellation task described in the text. The task is to strike through each line on a sheet of paper. Broken lines represent scenes. EM = eye movement. Numbers indicate head movement number. Letters indicate eye movement number within a single head movement.

The RPR does not *process information* or *compute space*; it uses a re-sizeable and moveable workspace centred on the fixation point – a *scene* – equivalent to the authors' FVF, but more precisely defined. It is an area within which every point is specified, by a set of coordinates, in terms of potential eye movements referenced to the fovea rather than physical distance. They become physical distance when the knowledge is instantiated.

The *scene* is used as follows (see Fig. 1):

1. The eyes fixate the centre of a line. Using the line object coordinates thus generated, a target line is cognitively (covertly) selected. The eyes then move to the target line, followed by the hand. Note, the *scene* – the set of coordinates – *moves with the eyes*, so that at each fixation point a new set of coordinates is generated. When the necessary eye movements have become

extended to a comfortable maximum, the head is shifted. The process is repeated until all of the lines have been cancelled.

2. The size of the *scene* is determined by the nature of the task, for example how precise it is, and how quickly it has to be performed. In driving a car a *scene* might be the entire visual field; in mending a watch, it might be just the fovea.

3. The choice of fixation point is again determined by the nature of the task. In this task, subjects tend to work from left to right and/or clockwise in a systematic fashion because this is natural and practical. In other cases the choice might be determined by colour or position.

The explanatory power of blending these cognitive data with neural data is illustrated by the work of Berti & Rizzolatti (2002), who have identified the special and distinct oculomotor and reaching circuits shown in Figure 2. In the oculomotor

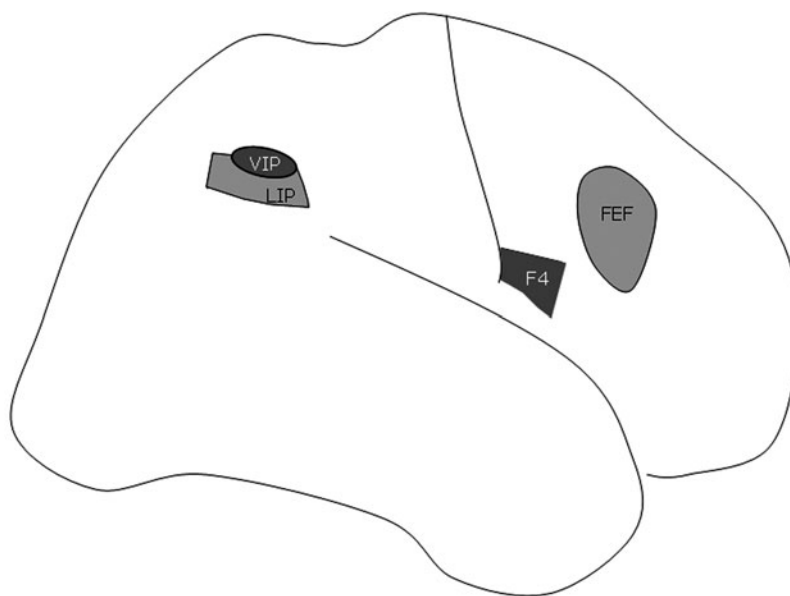


Figure 2 (Campion). Representation of macaque cerebral hemisphere. The light shaded area indicates the oculomotor circuit, and the dark shaded area the reaching circuit described in the text. LIP = Lateral intraparietal area. FEF = Frontal eye fields. VIP = ventral intraparietal area. F4 = pre-motor area.

circuit, neuronal receptive fields are insensitive to distance but are sensitive to stimuli falling only on a specific point on the retina referenced to the fovea, whereas neurons in the reaching circuit prefer three dimensional objects and have receptive fields coded in body part coordinates.

This explanation contrasts sharply with unhelpfully vague information processing terms such as parallel and serial processing used by the target article authors, and it does away with the concept of attention as a scientific term altogether. And we should note in passing that it is the strong dependency of cognitive operations on the nature of the task they subserve, coupled with the lack of explicit task analyses in research, that leads to the ambiguity and unresolved debate that characterises much cognitive psychology.

Brief though it is, I hope this illustration is sufficient to demonstrate the merits of the approach; I commend it to the authors and to the science.

Eye movements are an important part of the story, but not the whole story

doi:10.1017/S0140525X16000042, e134

Kyle R. Cave

University of Massachusetts, Amherst, Department of Psychological and Brain Sciences, Amherst, MA 01003.

kcave@psych.umass.edu

<http://people.umass.edu/kcave/>

Abstract: Some previous accounts of visual search have emphasized covert attention at the expense of eye movements, and others have focused on eye movements while ignoring covert attention. Both selection mechanisms are likely to contribute to many searches, and a full account of search will probably need to explain how the two interact to find visual targets.

Eye movements are an important part of many laboratory search tasks, and most real world search tasks. A complete account of visual search and visual attention will require an explanation of how eye movements are guided and how they contribute to selection. Furthermore, the ability to track eye movements has led to a succession of new insights into how attention is controlled. There are many examples, but one is Olivers et al.'s (2006) work investigating the working memory representations guiding search. Future eye tracking studies will almost certainly produce valuable new insights. Thus, in terms of both theory and method, eye movements have played a key role, and will continue to do so.

In the target article's framework, eye tracking data are combined with assumptions about the functional viewing field (FVF), the area within the visual field that is actively processed. The FVF is assumed to be small in difficult visual tasks, so that attentional processing during a single fixation is confined to a small region, and many fixations are necessary to search through a large array. The FVF can be expanded for easier tasks, allowing them to be completed with fewer fixations that are farther apart. As noted in the target article, the concept has been around for some time, but nonetheless it is difficult to demonstrate experimentally that the FVF is actually adjusted during search as Hulleman & Olivers (H&O) suggest; it cannot be measured as straightforwardly as tracking eye movements. However, H&O point out that evidence from gaze-contingent search experiments (Rayner & Fisher 1987; Young & Hulleman 2013) provide good evidence that information is being taken in from smaller regions in more difficult tasks. The FVF concept has also been useful in interpreting attentional phenomena other than search. For instance, Chen and Cave (2013; 2014; 2016) found that patterns of distractor interference that did not fit with perceptual load theory (Lavie 2005) or with dilution accounts (Tsal & Benoni 2010; Wilson et al. 2011) could be explained by assuming that

more difficult tasks induce subjects to adopt a narrower FVF (or as it is called in those studies, attentional zoom).

Thus, I agree on the importance of combining eye tracking data with assumptions about variations in FVF to build accounts of visual search. It is also clear that attentional theories need to be able to explain search in scenes that are not easily segmented into separate items. Earlier theories were clearly limited in these respects, in part because they were originally formulated at a time when it was more difficult to track eye movements. However, the proposed framework has other limitations. It appears to be built on the assumption that once the FVF size is set, there is nothing else for covert attention to do. That seems surprising, given the abundant evidence that covert attention can select locations based on color, shape, and other simple features within a fixation (Cave & Zimmerman 1997; Hoffman & Nelson 1981; Kim & Cave 1995). This selection can be done relatively quickly and efficiently (Mangun & Hillyard 1995). I am not trying to argue that attentional selection is fundamentally limited to one item at a time, but it is hard to believe that covert selection would not be employed during search to lower the processing load and limit interference within each fixation. In fact, shifts in covert attention can be tracked from one hemisphere to the other in the course of visual search (Woodman & Luck 1999). Given that covert attention can be adjusted more quickly than a saccade can be programmed and executed, it should be able to contribute substantially in investigating potential target regions and in choosing the next saccade, as suggested by a group of studies including Deubel and Schneider (1996), and Bichot et al. (2005).

Over the years, many attention researchers have tried to study visual search by focusing on covert attention and ignoring eye movements, while others have tried to focus on eye movements while ignoring covert attention. If the H&O framework is truly to be a hybrid approach, it seems that it should allow the possibility that many searches are accomplished through an interaction between eye movements and covert attention.

In considering the history of attention research, it is worth noting that the idea that attention can be adjusted between a broad distribution and a narrow focus has been explored in contexts other than Sanders' (1970) discussion of FVF mentioned in the target article. There is, of course, Eriksen and St. James' (1986) zoom lens analogy, but perhaps even more relevant for this discussion is Treisman and Gormican's discussion of how attention makes information about stimulus location available. Here is their description:

Attention selects a filled location within the master map and thereby temporarily restricts the activity from each feature map to the features that are linked to the selected location. The finer the grain of the scan, the more precise the localization and, as a consequence, the more accurately conjoined the features present in different maps will be. (Treisman & Gormican 1988, p. 17)

Although they do not explicitly refer to the functional field of view, it seems they had a similar concept in mind, as discussed in Cave (2012).

Another aspect of this framework is the move away from visual input that is organized into separate items. The motivation for this is clearly spelled out, but what is not explained is how the concept of object-based attention fits into this framework. There are some circumstances in which visual selection is apparently not shaped by the boundaries defining objects and groups (Chen 1998; Goldsmith & Yeary 2003; Shomstein & Yantis 2002), but they are rare, and the object organization of a display often affects attentional allocation even when it is not relevant to the task (Egley et al. 1994; Harms & Bundesen 1983). Is the claim in the target article that object and group boundaries play no role in visual search, even though their effects are difficult to avoid in other attentional tasks?

Searching for unity: Real-world versus item-based visual search in age-related eye disease

doi:10.1017/S0140525X16000054, e135

David P. Crabb and Deanna J. Taylor

Division of Optometry and Visual Science, School of Health Sciences, City, University of London, London EC1V 0HB, United Kingdom.

david.crabb.1@city.ac.uk

<http://www.staff.city.ac.uk/crabblab>

deanna.taylor@city.ac.uk

<http://www.staff.city.ac.uk/crabblab>

Abstract: When studying visual search, item-based approaches using synthetic targets and distractors limit the real-world applicability of results. Everyday visual search can be impaired in patients with common eye diseases like glaucoma and age-related macular degeneration. We highlight some results in the literature that suggest assessment of real-world search tasks in these patients could be clinically useful.

Visual search is an important everyday visual function. Many laboratory studies of visual search use synthetic targets and distractors. Such experiments are somewhat removed from the holistic approach needed to find a face in a crowd, search for an exit sign at an airport or locate a favourite cereal on the supermarket shelf. So we agree with Hulleman & Olivers' (H&O's) contention that "item-based approaches limit the real-world applicability of results from the lab" (sect. 4.1, para. 1). H&O discuss two real-world applications of visual search: radiology and airport security. In this commentary we highlight search as an impaired everyday visual function in people with age-related eye disease. We speculate on how this might be best assessed with the idea of bringing visual search out of the lab and into clinical research, focussing on open angle glaucoma and age-related macular degeneration (AMD), two of the most common causes of visual impairment worldwide (Lamoureux et al. 2008). Glaucoma is typically associated with peripheral vision loss, whilst AMD causes loss of central vision.

Most studies of visual search in age-related eye disease used an item-based approach (for example, Jacko et al. 2000; 2001), yet examples taking a more real-world approach are emerging. We investigated visual search in people with glaucoma using two computer-based tasks (Smith et al. 2011), one item-based task requiring participants to identify a target from an array of distractors, and another more real-world task requiring participants to find everyday items in digital photographs of indoor and outdoor scenes. Participants with glaucoma exhibited longer average search times than healthy peers for the real-world task, whilst search times were not significantly different between the two groups for the item-based task. These results support the notion that item-based search tasks are not relatable to real-world applications. A further study (Smith et al. 2012), investigating eye movements during the same real-world visual search task, reported a reduction in saccade frequency in people with glaucoma compared with healthy peers. Furthermore, amongst participants with glaucoma, those who made more saccades per second were quicker in finding the real-world targets. These results indicate that eye-movement behaviour is of importance when considering visual search performance of people with age-related eye disease, and were supported by a study of similar design when detecting faces (Glen et al. 2013). These findings align with H&O's proposition that fixation count is a critical factor in visual search behaviour.

Fixation count has been investigated in real-world search tasks in AMD. Most visual search research in AMD has been conducted using artificial arrays (for example, searching for a letter T amongst distractors in the form of the letter L) and participants with simulated scotomas (for example Bertera 1988; Coeckelbergh et al. 2002; Cornelissen et al. 2005; Geringswald et al. 2012; 2013; Kuyk et al. 2005; MacKeben & Fletcher 2011; Murphy & Foley-Fisher 1988; 1989). These approaches allow for more controlled experimental design, yet simulated scotomas

may not be entirely realistic (Harvey & Walker 2014; Schuchard et al. 1999). One method of simulating central scotoma uses contact lenses with a central opacity, which cause reduced retinal illumination, leading to worsening in visual acuity and contrast sensitivity (Butt et al. 2015). A gaze-contingent simulation of scotoma, incorporating eye tracking, is likely to provide better scotoma simulation (Butt et al. 2015); we have used this in a hazard search task in driving (Glen et al. 2015). Results were useful, but simulation cannot capture the real experience of patients, where self-reported perception and description of scotoma varies enormously (Crabb et al. 2013). A few studies have investigated real-world visual search in actual patients with AMD; for example Thibaut et al. (2015) reported individuals with AMD exhibit higher saccade frequencies, shorter fixation durations and longer scan paths compared with those without AMD during visual search. Aspinall et al. (2014) found fixation count to be a useful marker of situations subjectively classed as "difficult" by individuals with AMD when assessing eye movement behaviour whilst watching videos of ambulatory journeys. Similarly, Geruschat et al. (2006) investigated gaze behaviour during street crossing and reported higher fixation count during more difficult/visually demanding parts of the task. Seiple et al. (2013) observed people with AMD whilst exploring faces and reported fixation count for internal facial features (eyes, nose, and mouth) to be higher for controls than for individuals with AMD. All of these tasks transcend the traditional item-based search.

Studies of everyday visual search have real clinical implications. Visual search in people with visual impairment has been suggested as a predictor for mobility and performance of other daily activities (Kuyk et al. 2005). There is evidence for the effectiveness of eye movement training on visual search in congenital prosopagnosia (Schmalzl et al. 2008), following brain damage (Bouwmeester et al. 2007), and for improved visual search and mobility performance in people with visual impairment of ocular origins following repeated practice of an item-based search task (Kuyk et al. 2010; Liu et al. 2007). These types of findings could lead to interventions and alternative approaches to management of patients. Potential also exists for development of tests for detecting and monitoring eye disease by using visual search in both item-based (Loughman et al. 2007) and real-world (Crabb et al. 2014) scenarios.

An article published nearly 30 years ago about tumour detection using visual search (Nodine & Kundel 1987) states that "detecting an object that is hidden in a natural scene is not the same as detecting an object displayed against a background of random noise." Research in this area ought to bridge the gap between lab-based testing and the real world. H&O have made an important step towards unifying some of the theory of visual search. We anticipate this will stimulate practical studies that may lead to better understanding of visual search in people with age-related eye disease. In turn we speculate that this will have implications for rehabilitation, and potentially lead to development of new tests for monitoring age-related eye disease.

"Target-absent" decisions in cancer nodule detection are more efficient than "target-present" decisions!

doi:10.1017/S0140525X16000066, e136

Trevor J. Crawford,^a Damien Litchfield,^b and Tim Donovan^c

^aDementia Research and Eye Tracking Lab, Faculty of Science and Technology, Psychology Department, Lancaster University, Lancaster LA1 4YF, United Kingdom; ^bDepartment of Psychology, Edge Hill University, Ormskirk L39 4QP, United Kingdom; ^cMedical & Sport Sciences, University of Cumbria, Lancaster LA1 3JD, United Kingdom.

t.crawford@lancaster.ac.uk tim.donovan@cumbria.ac.uk
litchfld@edgehill.ac.uk

Table 1 (Crawford et al.). Mean number of fixations per zone ($n = 27$ X-ray films)

Chest Zone Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total	SD
Radiologists	16.3	18	19.5	20.1	15.6	16	18.5	19.9	19.6	14.3	13.5	18	18	17.3	245	2.09
Radiographers Post-training	26.4	23.6	25	32.2	29.2	22.6	31	29	30.2	26.8	19.6	31.6	31.4	29.6	388	3.84
Radiographers Pre-training	31.8	28.6	30	32.2	29.5	22.6	30	29	30	26.8	19.6	31	32	29.8	403	3.6
Novices	30.3	28.5	29	29.8	29.6	23	30	30	31	27	20	30	30	31	399	3.18

Data from Manning et al. (2006).

[http://www.research.lancs.ac.uk/portal/en/people/trevor-crawford\(55e99f0c-4f92-437e-9a68-60d7f94265f2\).html](http://www.research.lancs.ac.uk/portal/en/people/trevor-crawford(55e99f0c-4f92-437e-9a68-60d7f94265f2).html)
<https://www.edgehill.ac.uk/psychology/dr-damien-litchfield/>
<https://www.cumbria.ac.uk/study/academic-staff/all-staff-members/tim-donovan-phd.php>

Abstract: Many parts of the medical image are never fixated when a radiologist searches for cancer nodules. Experts are able to use peripheral vision very efficiently. The size of the functional visual field appears to increase according to the level of expertise. However, searching a medical image diverges, in a puzzling way, from the typical search for a target feature in the laboratory.

There has been little change in the proportion of medical decision errors in radiology over the last 60 years, despite substantial advances in technology. The field has not succeeded in capturing or understanding the fundamental properties of visual search and the allocation of visual attention of the expert radiologist, nor in translating the essential search skills into training programs. We therefore welcome this work, in the hope that a new bridge will be developed that will connect visual science with this radiological challenge. As Hulleman & Olivers (H&O) state, the fields of medical imaging visual search “have been underserved by item-based models (or any form of overarching theory of search...)” (sect. 6.6, para. 7). What constitutes an “item” in the medical image is not at all obvious. H&O suggest that the focus on the individual search items should now give way to a greater emphasis the properties of the functional visual field (FVF).

There is some evidence from our own work that this approach is relevant to visual search with medical images. In tasks where we have conducted eye-tracking experiments on groups with differing levels of expertise, we found that experts will typically make fewer fixations than novice observers. This happens for a relatively straightforward task such as fracture detection in bones (Donovan et al. 2005) and also in more complicated tasks, such as chest radiographs with many potential “items” or structures which resemble pathology. We have demonstrated distinct differences between radiologists (experts), radiographers (pre- and posttraining in chest radiograph interpretation) and novice observers when searching for lung nodules in chest radiographs. Experts find many more lung nodules while generating fewer fixations and larger saccadic amplitudes (see Manning et al. 2006). This supports the idea that the FVF is modifiable and does change according to the level of expertise. The work also sheds some light on the timescale of this learning or plasticity. After six months of training the number of fixations of the radiographers had decreased compared with their pre-training levels but had not reached that of the expert radiologist (see Table 1). Importantly, as well as making fewer fixations there was a more uniform distribution of fixations across all regions of the chest radiograph by the experts, suggesting that once the FVF has adapted to the task as a result of training, it is applied consistently across the medical image.

More direct evidence of modifications to FVF could be explored with gaze-contingent display paradigms to isolate the expertise-dependent changes in visual search from the benefits (and costs) of initially processing the entire scene (Litchfield & Donovan 2016). The link of fixations with the speed of RTs points to one of the hallmarks of expertise—that experts are able to find targets

faster and with fewer fixations than novices (Reingold & Sheridan 2011). Another hallmark of expertise, which is best confirmed using gaze-contingent paradigms, is that the perceptual span increases as a function of expertise (Charness et al. 2001; Kundel et al. 1984; Rayner 2009). Simple RT slopes have not helped us to understand why so many cancers are missed in medical imaging; therefore, we appreciate the central role of the FVF in this conceptual model. Unpacking the dynamic nature of FVF as a function of task and expertise may yield greater insight into this process.

However, we see an area of concern: The model replicates the conventional finding that target-present decisions are conducted more quickly than target-absent decisions (in medical images, “target-absent” would be equivalent to the true negative images – i.e., images where no cancer modules are present). H&O state “...all models of visual search, including the framework presented here, seem much better at describing target-present trials than target-absent trials” (sect. 6.6, para. 6). In our study of chest X-rays where some films showed cancerous nodules and some did not, the target-absent decisions (true negatives) were *faster* than the true positive decisions (see Manning et al. 2005). Interestingly, this applied to both the experts (radiologists) and the novices. Our concern therefore goes beyond the lack of an explicit stop search signal. There appears to be a fundamental reversal in the normal pattern of target-absent versus target-present decisions when visual search is conducted with a chest X-ray.

Recently, Litchfield & Donovan (2016) used a gaze-contingent preview to explore the effects a preview window in the domain of a naturalistic scene versus a medical image for radiologists and novices. The work found a clear dissociation between the two domains, with a strong preview benefit on the visual search performance for naturalist scenes, but no benefit with medical images for either group. Thus, our earlier and more recent work urges caution in extrapolating across the different search domains of feature search tasks, naturalist scenes and medical images. This suggests the bridge that the authors are seeking to construct will be more complex than they envisaged.

Why the item will remain the unit of attentional selection in visual search

doi:10.1017/S0140525X16000078, e137

Martin Eimer

Department of Psychological Sciences, Birkbeck, University of London, London WC1E 7HX, United Kingdom.

m.eimer@bbk.ac.uk

<http://brainb.psyc.bbk.ac.uk/>

Abstract: Hulleman & Olivers (H&O) reject item-based serial models of visual search, and they suggest that items are processed equally and globally during each fixation period. However, neuroscientific studies have shown that attentional biases can emerge in parallel but in a spatially selective item-based fashion. Even within a parallel architecture for visual search, the item remains the critical unit of selection.

Leading theories of visual search postulate that search targets are found by deploying attention sequentially to individual objects (items). Hulleman & Olivers (H&O) reject such serial item-based accounts and propose an alternative where fixations replace items as the conceptual unit of visual search. In their nascent computational model, individual search episodes start once the eyes have reached a new fixation location. Parallel processing of all items within a functional view field (FVF) results in a decision about target presence/absence. If no target is found, the eyes move to a different location, and a new search episode commences. This model performs remarkably well in simulating search slopes and the variability of search performance across different types of search tasks. However, questions remain about the mechanisms proposed for localizing targets and discriminating them from irrelevant objects during individual fixations. For example, fixation duration is constant at 250 ms, and the visual slate is wiped clean during each new eye movement, and therefore the decision about the presence of a target within the FVF has to be made within this brief time window. Results from attentional dwell time and attentional blink experiments suggest that target identification processes may require at least 300–500 ms, and may therefore extend in time beyond individual fixation periods.

At a more fundamental level, it is difficult to see how objects can be replaced as conceptual units in visual search, given that the visual world is made up of objects, and finding a particular target object is the goal of a typical search task. H&O claim that processing with a fixation period is not item-based, because “all items are in principle selected and processed simultaneously” (sect. 6.3) by mechanisms that compute global area activations and pooled summary statistics across the FVF. This is plausible for easy search tasks where targets can be found on the basis of local feature discontinuities (singleton detection), and also for non-search tasks that require the rapid extraction of the gist of a scene. What remains unclear is whether such global area-based mechanisms can detect the presence or absence of targets even in moderately difficult search tasks where no diagnostic low-level saliency signals are available and distractors share features with the target. Furthermore, the spatially non-selective group-based account proposed by H&O seems at odds with neuroscientific insights into the control of visual search. During search for targets with known features, biases of visual processing towards target-matching objects emerge rapidly within the first 200 ms after the presentation of a search display, even outside of the current attentional focus (e.g., Bichot et al. 2005). These biases are elicited in a spatially specific fashion in retinotopic visual areas that match the location of possible target objects. They can initially be triggered at multiple locations across the visual field, but gradually become more spatially focused, and may eventually result in the selective activation of one particular object representation (see Eimer 2014; 2015, for a more detailed discussion, and Duncan 2006, for related ideas on object-based integrated competition mechanisms in visual search). The important point here is that such task-dependent attentional biases of visual processing emerge in spatial visual maps that represent candidate target objects at particular locations. In this fundamental sense, attentional selection mechanisms and their neural basis remain irreducibly item-based. Crucially, this type of item-based selectivity does not imply serial selection. Spatially selective processing biases for target-matching objects can emerge in parallel across the visual field (e.g., Bichot et al. 2005; Saenz et al. 2002), and multiple target objects at different locations can be selected simultaneously and independently (e.g., Eimer & Grubert 2014).

Within the framework proposed by H&O, it may be useful to distinguish between the guidance of spatial attention during individual fixation episodes, and the guidance of eye movements. The selection of new fixation locations might indeed be informed by global area-based computations that are performed in parallel outside of the currently fixated region, and provide information about the likelihood of a target being present elsewhere in the visual field. In contrast, attentional control processes within the FVF during a fixation

episode operate via spatially selective and thus essentially item-based modulations of visual processing. In fact, H&O acknowledge the existence of such spatial biases that gradually become more item-based for the case of compound search where target-defining and response-relevant features differ. Here, “a global search for the target-defining feature may be followed by a local search for the response-defining feature.” The question remains whether this type of item-based spatially selective attentional control is the exception or the rule during visual search. Although some real-world visual search tasks (e.g., the scanning of mammograms or security X-ray images) do not involve the well-defined objects that are used in lab-based search studies, one could argue that even here, search is still guided in a spatially selective fashion by image features that are relevant for the task at hand.

The new fixation-based search model proposed by H&O is useful not only because of its power to simulate behavioural results, but also because it invites us to think differently about visual search. Serial selection models have dominated the field for decades, and alternative concepts are sorely needed. H&O provide excellent arguments for abandoning strictly sequential item-by-item accounts of visual search. However, in their endeavour to reject serial selection, they may have thrown out the item-based baby with the serial bathwater. Attentional processes in visual search may indeed operate in a largely parallel fashion, but the item will remain a primary unit of selection.

Fixations are not all created equal: An objection to mindless visual search

doi:10.1017/S0140525X1600008X, e138

James T. Enns^a and Marcus R. Watson^b

^aDepartment of Psychology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; ^bDepartment of Biology, York University, Toronto, ON M3J 1P3, Canada.

jenns@psych.ubc.ca
http://visionlab.psych.ubc.ca
marcuswatson@gmail.com

Abstract: This call to revolution in theories of visual search does not go far enough. Treating fixations as uniform is an oversimplification that obscures the critical role of the mind. We remind readers that what happens during a fixation depends on mindset, as shown in studies of search strategy and of humans' ability to rapidly resume search following an interruption.

We welcome Hulleman & Olivers' (H&O's) invitation to abandon the display item as the fundamental unit of visual search. There is now considerable evidence – some of which we have contributed to (Enns & Kingstone 1995; Fecteau et al. 2000; Roggeveen et al. 2004; van Zoest et al. 2006) – that display items cannot be considered in isolation from the items around them, nor from the limits of the observer's functional viewing field (FVF). However, H&O's call to revolution does not go far enough because they simply replace one operational unit (the experimenter-defined item in a search display) with another (the observer's FVF, as indexed by fixations). Both of these efforts to ground theories of search in easily observable third-person variables neglect the most important factor: the observer's *mind*. It is our view that what happens behind the observer's eyes is more important than what happens in front of them (the display items) or even in them (the FVF).

Fixations during visual search cannot be considered in isolation; they are always involved in a trading relationship with saccades. That is, at any given moment the observer is engaged in strategic decisions (albeit implicit ones) to keep their eyes still (allowing for *seeing*, the ability to distinguish targets from non-targets) or to move them (allowing for *looking*, the acquisition of new information from outside of the current fixation). Similar to agents in classical reinforcement learning models (Sutton & Barto 1998), who

trade off between *exploiting* currently available resources and *exploring* for novel resources, visual searchers cycle between *seeing* the information in their current fixation and *looking* to a new location. In typical search tasks this cycle is repeated 3–4 times per second, which is consistent with H&O's decision to model fixations as lasting 250 ms. Yet setting this average time as a constant conceals important variability.

We recently manipulated the mental strategies of participants by randomly assigning them to either search *passively*, by giving them instructions to “let the target pop into your mind,” or to search *actively*, by telling them to “deliberately direct your attention.” We found a passive response time advantage, as in previous studies (Smilek et al. 2006), and showed that this stemmed from the inherent trade-off between seeing and looking (Watson et al. 2010). Passively instructed participants made fewer fixations of longer duration, and once they fixated the target region for the first time, they made fewer subsequent fixations, responding more quickly. This suggests that they placed a greater emphasis on *seeing* rather than looking, and so were better able to process the target. These differences may not reflect differences in FVF: passively instructed participants were no more likely to fixate closer to the center of the display, nor further away from individual items, either of which would have allowed them to take advantage of a larger FVF.

One of the most striking observations in Watson et al. (2010) is that there is more than one way to succeed in visual search. Even after trimming participants from each group to equate overall speed and accuracy, passively instructed participants made fewer fixations separated by larger saccades. This means it is possible to trade the higher information resolution of *seeing* with the greater information acquisition of *looking*, without affecting overall success.

Another demonstration of the critical importance of mind comes from studies of *rapid resumption* of visual search, in which participants are able to accurately respond very rapidly (within 100–400 ms) to a display that has been re-presented following a brief interruption (Lleras et al. 2005). To be as accurate, the same responses to the first look at a display take more than 500 ms. Successful rapid resumption of an interrupted search depends on participant's forming a mental prediction of the target's response-relevant features and location based on a first look at the display. When these features or location change after the first look, but all other aspects of the display remain constant, rapid resumption is eliminated (Lleras et al. 2007). The prediction is more likely to be made when fixations are located close to the target, but it turns out that fixation location is not the determining factor either. When gaze-contingent displays are used, such that the target is always presented at fixation, rapid resumption is impossible (van Zoest et al. 2007). Once again, differences in the mind lead to differences in the processing that occurs during a fixation. If a correct prediction about the target has been made during the first glance, the second glance enables rapid responses; if this prediction has not been made, the search must be started over. This critical *predictive* aspect of visual search seems absent from H&O's account.

The role of fixations depends on mind-set, both between-subjects (as shown by our instructional study) and within-subjects (as shown by rapid resumption studies). H&O's account treats fixations as uniform, which is a serious oversimplification.

We conclude by reiterating three points Hochberg (1968) made long ago, recently updated in a series of studies in Peterson et al. (2006). First, because every percept enters the mind through piecemeal views, a theory of perception must be about the mind's representation, not its trigger (the stimulus), nor its conduit (the eye). Second, as Hochberg liked to put it, “unlike objects themselves, our perception of objects is not everywhere dense” (Hochberg 1982, p. 214). He contributed numerous demonstrations of the selectivity of perception, both in its overt actions (fixations) and in its covert processing (attention). Third, there is no perception in a glance that is divorced from prior mental representation. This recursive aspect of perception means that vision

is as much influenced by what lies in the mind as what lies in the eye of the beholder.

Contextual and social cues may dominate natural visual search

doi:10.1017/S0140525X16000091, e139

Linda Henriksson^a and Riitta Hari^{a,b}

^aDepartment of Neuroscience and Biomedical Engineering, Aalto University, 00076 AALTO, Finland; ^bDepartment of Art, Aalto University, 00076 AALTO, Finland.

linda.henriksson@aalto.fi riitta.hari@aalto.fi

https://people.aalto.fi/linda_henriksson

https://people.aalto.fi/riitta_hari

Abstract: A framework where only the size of the functional visual field of fixations can vary is hardly able to explain natural visual-search behavior. In real-world search tasks, context guides eye movements, and task-irrelevant social stimuli may capture the gaze.

Can visual search be explained by a model with only one free parameter, the size of the functional visual field (FVF) of a fixation, as suggested by Hulleman & Olivers (H&O)? Considering fixations, rather than individual items, as the primary unit of visual search agrees with the tight connection between eye gaze and information retrieval. H&O demonstrate that their framework successfully captures the variability of reaction times in easy, medium and difficult searches of elementary visual features. However, beyond laboratory conditions (“find a specific item among very similar distractors”), visual search strategies can hardly be explained by such a simple model because the search space is poorly specified (e.g., “Where did I leave my keys?”, “Is my friend already here?”), and the search strategy is affected, for example, by experience, task, memory, and motives. Moreover, some parts of the scene may attract attention and eye-gaze automatically because of their social and not only visual saliency.

In real-life situations, the search targets are not a priori evenly distributed in the visual field, and the task given to the subject will affect the eye movements (Neider & Zelinsky 2006; Torralba et al. 2006; Yarbus 1967). Moreover, the scene context can provide spatial constraints on the most likely locations of the target(s) within the scene (Neider & Zelinsky 2006; Torralba et al. 2006). The viewing strategy is also affected by expertise: experienced radiologists find abnormalities in mammography images more efficiently than do less-experienced colleagues (Kundel et al. 2007); experts in art history and laypersons view paintings differently (Pihko et al. 2011); and dog experts view interacting dogs differently than do naïve observers (Kujala et al. 2012). Moreover, the fixation durations vary depending on the task and scene: Although all fixations may be of about the same duration for homogeneous search displays, short fixations associated with long saccades occur while exploring the general features of a natural scene (ambient processing mode) and long fixations with short saccades take place while examining the focus of interest (focal processing mode; Unema et al. 2005).

H&O suggest that the concept of FVF would allow semantic biases in visual search by accommodating multiple parallel FVFs—for example, a small FVF for the target object and a larger FVF for recognizing the scene. This extension might account for processing within the fixated area, but could it also predict saccade guidance? Predicting eye movements occurring in the real world would require a comprehensive model of the semantic saliency of the scene, which is really challenging. That said, the recent advances in neural network modeling of artificial visual object recognition (Krizhevsky et al. 2012) could facilitate the modeling of the semantic and contextual features that guide the gaze (Kümmerer et al. 2014).

Finally and importantly, social cues strongly affect natural visual processing. Faces and other social stimuli efficiently attract gaze (Birmingham et al. 2008; Yarbus 1967), insofar as a saccade toward a face can be difficult to suppress (Cerf et al. 2009; Crouzet et al. 2010). Thus, the mere presence of a task-irrelevant face can disrupt visual search by attracting more frequent and longer fixations than do other distractors (Devue et al. 2012). Such a viewing behavior contrasts with the conventional search tasks that become more difficult when the resemblance of the distractors and target increases. Whereas faces capture attention (and gaze) in healthy subjects, autistic individuals are less distracted by social stimuli in the search scene (Riby et al. 2012) and experience reduced saliency in semantic-level features, especially in faces and social gaze, during free-viewing of natural scenes (Wang et al. 2015). Altogether, social stimuli have such a central role in human behavior and brain function (Hari et al. 2015) that they should not be neglected in models aimed to explain natural visual-search behavior. Peripheral vision can provide effective summary statistics of the global features of the visual field (Rosenholtz 2016), and thus social stimuli, such as faces, outside of the foveal vision could significantly affect the visual search.

Face recognition represents a special case of visual search—a natural search task could be, for example, to find a friend among a crowd of people. For (Western) faces, the optimal fixation location is just below the eyes (Peterson & Eckstein 2012), and two fixations can be enough for face recognition (Hsiao & Cottrell 2008) for isolated face images. Whether the same is true for faces in their natural context remains to be seen. Overall, it appears that the saccades to faces and to scenes are consistent across subjects during the initial viewing and become less consistent during later saccades (Castelhano & Henderson 2008). In addition, the initial saccades are consistent across cultures, with saccade endpoints reflecting the optimal fixation locations in face identification tasks (Or et al. 2015). These findings raise interesting questions related to the neural underpinnings of natural visual search: How does the guidance of the initial saccades differ from later saccades? At what level of cortical processing does the cultural background or expertise affect the saccade guidance?

In conclusion, we doubt that “an overarching framework of visual search” can be built without implementing effects of contextual and social cues. Building a model that can predict an observer’s eye movements during natural search tasks in real-world visual environment remains a challenge.

Until the demise of the functional field of view

doi:10.1017/S0140525X16000108, e140

Laurent Itti

Computer Science Department, Psychology Department and Neuroscience Graduate Program, University of Southern California, HNB-07A, Los Angeles, CA 90089-2520.

itti@usc.edu

http://iLab.usc.edu

Abstract: Hulleman & Olivers (H&O) make a much-needed stride forward for a better understanding of visual search behavior by rejecting theories based on discrete stimulus items. I propose that the framework could be further enhanced by clearly delineating distinct mechanisms for attention guidance, selection, and enhancement during visual search, instead of conflating them into a single functional field of view.

The target article presents a laudable effort to depart from conceptual analyses of visual search that are based primarily on discrete stimulus items. By using fixations as the central unit, the authors propose a significant paradigm shift for core research on visual search. As the authors note, this idea has been proposed before. In fact, it is central to most computational theories and

models of visual attention. Indeed, the standard test for saliency map models consists of measuring their ability to predict human fixation locations in a wide range of tasks (Borji et al. 2013), including visual search (e.g., Ehinger et al. 2009). A framework based on fixations promises an understanding of visual search behavior that is more focused on perceptual and motor mechanisms than on external stimulus composition.

Computational models of attention—applicable not only to visual search but also to virtually any visually guided behavioral task—often distinguish among at least three facets of visual attention (Itti & Borji 2013):

1. *Guidance:* What computations are involved in deciding where or what in a scene to attend to next? This includes bottom-up or stimulus-driven guidance signals, such as visual saliency, and top-down guidance signals, for example related to spatial or feature-based cueing.

2. *Selection:* How is a fraction of the visual input separated out from other incoming visual signals so that it can be processed in more detail?

3. *Enhancement:* How are selected (attended to) visual items processed differently—usually better by some measure such as enhanced contrast sensitivity or better discrimination threshold—than nonselected (unattended) items?

All three aspects have been studied extensively in the electrophysiology, psychophysics, and modeling literature (for reviews, see, e.g., Allport et al. 1993; Borji & Itti 2013; Carrasco 2011; Desimone & Duncan 1995; Driver & Frith 2000; Itti & Koch 2001; Reynolds & Desimone 1999; Robertson 2003). The target article authors deliberately sidestep guidance in their simulations and focus on selection, in particular through the concept of functional field of view (FVF), which is akin to an attention spotlight (Crick 1984), and they omit enhancement.

The framework could be strengthened by clearly embracing the idea of separate functional mechanisms (although possibly overlapping and using shared neural components) for guidance, selection and enhancement. Indeed, with a single FVF, the concepts of scene gist and contextual guidance already seem to be a struggle for the framework (sect. 6.2): Presumably, when starting a new search over a natural scene, one would first need a wide FVF for scene identification and to establish semantic guidance, then switch to a smaller FVF during search. Thus, as the authors concede, the FVF size is unlikely to be fixed as hypothesized in the simulations, and may instead change before and possibly during search. Even with a dynamic FVF, it is unclear whether phenomena such as contextual guidance or cueing of search—in natural or artificial scenes (Chun & Jiang 1998; Torralba et al. 2006)—would require the same FVF processing characteristics as would be required to select and analyze items at the current fixation. An FVF that can rapidly change size and functional form presents little conceptual value as it is an unknown time-varying entity. It becomes a liability for the framework because any unexplained phenomenon could be attributed to some sudden change in FVF size or processing form that one would be hard-pressed to measure in real time.

Thus, one may need to split the concept of FVF into at least two or three: possibly a broader FVF operating coarser or more statistical processing might be necessary to compute a saliency map around the current fixation and to integrate contextual information broadly, to provide effective *guidance* of search. But a narrower and more selective FVF may sometimes be needed for *selection*, for example in difficult compound search. *Enhancement* may require a different FVF as well, possibly even more tightly wrapped around the selected items to reject background noise within the selected region. All three might be required in complex search scenarios (e.g., compound search on a noisy yet

semantically meaningful natural scene background). Positing a single FVF unnecessarily conflates guidance, selection and enhancement. This is already evident in the second component of the framework (sect. 5.2), where a target may be easy to detect peripherally because of a distinctive feature but hard to identify because of a more complex response-defining attribute. The authors assume that a broad FVF should be deployed for guidance of each fixation, followed by a switch to a narrow FVF for selection and identification of items. As with contextual guidance, this dynamic resizing essentially amounts to interlacing over time two different FVFs, one for guidance and the other for selection. Because in addition to different sizes (which may themselves be dynamic), these two FVFs might also have different processing characteristics (e.g., with respect to crowding or lateral masking), I suggest that decoupling the two into a guidance mechanism and a selection mechanism might enhance the long-term usefulness and robustness of the framework (and likewise for an enhancement mechanism).

Electrophysiological and neuroimaging evidence supports, at least in part, distinct mechanisms for guidance, selection, and enhancement. In primates, including humans, guidance likely recruits saliency mechanisms in the dorsal visual processing stream (e.g., Itti & Koch 2001; Kusumoki et al. 2000), and scene-level contextual mechanisms, for example in the lateral occipital complex (area LOC; e.g., Preston et al. 2013). The frontal eye fields (FEF) and ventral prearcuate (VPA) areas of the prefrontal cortex are thought to provide goal-directed spatial and feature-based modulation of guidance (Bichot et al. 2015). Shrink-wrapping of receptive fields in the ventral processing stream including inferotemporal cortex may serve selection and enhancement of selected information (Desimone & Duncan 1995; Zhang et al. 2011).

Although the authors of the target article wish to remain at a conceptual rather than algorithmic level, I believe that the best route to avoiding possible conceptual ambiguities or conflating between what should be distinct entities will have to be through detailed computational implementation of the components of this exciting framework.

ACKNOWLEDGMENTS

Professor Itti's research is supported by the National Science Foundation (grant numbers CCF-1317433 and CNS-1545089), the Army Research Office (W911NF-12-1-0433), and the Office of Naval Research (N00014-13-1-0563). The author affirms that the views expressed herein are solely his own and do not represent the views of the United States government or any agency thereof.

Feature integration, attention, and fixations during visual search

doi:10.1017/S0140525X1600011X, e141

Abbas Khani^{a,b} and Mehdi Ordikhani-Seyedlar^c

^aFunctional Brain Mapping Laboratory, Department of Basic Neurosciences, University of Geneva, 1211 Geneva, Switzerland; ^bVisual Cognition Laboratory, Department of Medicine, University of Fribourg, 1700 Fribourg, Switzerland; ^cDepartment of Neurobiology, Duke University, Durham, NC 27710.

abbas.khani@unige.ch mehdi.ordikhani@gmail.com

Abstract: We argue that mechanistic premises of “item-based” theories are not invalidated by the fixation-based approach. We use item-based theories to propose an account that does not advocate strict serial item processing and integrates fixations. The main focus of this account is feature integration within fixations. We also suggest that perceptual load determines the size of the fixations.

The target article reviews a number of previous theories of visual search and argues that a better understanding of visual search is

hampered by the assumption that individual items are central units of visual search. To explain visual search behavior (VSB), Hulleman & Olivers (H&O) propose a compelling alternative framework in which fixations are central units of the VSB and suggest that the demise of the item in visual search is impending. We greatly appreciate this target article drawing attention to the importance of fixations in visual search and to other ways of conceiving visual search behavior, but we object to the proposition that the item and item-based theories in explaining VSB are flawed. In our opinion, whereas fixations are a necessary element of the VSB, they are not sufficient for an integrative account of visual search. We argue that both peripheral and central processes contribute to the VSB and therefore an integrative account of the VSB will include fixations as well as elements of the established theories of VSB such as feature integration and attention. The authors acknowledge that the theories they call “item-based” recognize the existence of large amounts of parallel processing and that some of these theories are not based on individual items. Therefore, it is possible to consider these theories in ways other than strict serial processing of items. H&O have claimed that “within fixations, items are processed in parallel.” We reconsider this by highlighting the role of attention in visual search. The obligatory relationship between eye movement and attentional shifts in which eye movement cannot be performed without the attentional shifts has long been identified (Fischer 1987). Fixations occur to cluster items together. We agree with the authors that subjects tend to move their eyes because “covert search is much harder” (sect. 6.4, para. 2). However, we emphasize that within each fixation, covert attention plays a critical role on serial processing of individual items (Buschman & Miller 2009). In a recent study Marti et al. (2015) used a unique strategy in which subjects had to report their fixations in a search task. The results of self-reports were then compared with the actual eye movements. They showed that in some cases, subjects reported eye movements that they had never made. They concluded that item search was conducted by covert attention strategy and they had probably reported covert shifts of attention as eye movements. This indicates the importance of items in search strategy within each fixation.

Regarding feature integration theory (FIT; Treisman & Gelade 1980), feature integration and fixations are reconcilable in our proposal. H&O invalidate FIT as a viable account of VSB because this theory has classically been used to advocate serial processing of items arising from the conjunctions of different features. Although conjunctions are necessary for full perception, it is not necessary to perceive full conjunctions with a full map of features that lead to serial processing of the items. Feature extraction takes place at several levels and it does not need complete scrutiny at every level as there is evidence that humans can recognize degraded images such as faces (Gilad-Gutnick & Sinha 2012). In our account, at the first fixation, incomplete feature maps are made which gives a gist of the whole scene. These maps are made randomly, though the most salient features (Xiaodi et al. 2012) have a higher chance to enter these maps. Rather than conjunctions that lead to a full perception of individual items, loose conjunctions and clusters of similarities among features are made (Oliva & Torralba 2006). Using these maps, parallel exclusions and inclusions guide attention covertly or overtly to the most informative areas of the visual scene. This guidance leads to a more detailed map. At this level, more detailed (though not necessarily complete) parallel feature maps are formed within each fixation. Whenever an item or a number of items passes a certain threshold of similarity with the template, those individual items might be examined serially within the fixation, which can lead to either a target-present response or continuation of the search task. This is specifically true in the case of real world situations such as searching for a lesion in a radiographic image.

An important question is the size of each fixation or functional viewing field (FVF). The extent of feature extraction/integration

depends on the size of the FVF. H&O argue that the difficulty of discriminating items determines the size of the FVF. We propose that the fixation size is determined by the perceptual load. Following earlier work of Kahneman and Treisman (1984), Lavie demonstrated that perceptual load is the major determinant of the locus of selection in visual attention (Lavie & Tsai 1994) and that perceptual load is necessary for early selection (Lavie 1995). According to the load theory of attention, the scope of perception will be stretched from the center of the fixation to the surrounding area to the extent that the perception is loaded. It has to be noted that although this is a theory of attention, unlike cognitive load, perceptual load involves the early sensory system. To enable feature integration, the size of the fixations is adjusted according to the perceptual load of a group of items. In larger FVFs (e.g., initial fixation), the perceptual load is saturated with incomplete feature extraction. In our account, fixations are a measure, not central unit, of the feature integration at different levels.

In conclusion, H&O present a powerful case to support a framework that unifies fixation-based studies of VSB. However, their RT-based arguments to invalidate item-based theories of VSB need to be revisited. We argue that perceptual load determines the size of the fixations and consequently the number of the fixations. In a step towards an integrative account of the VSB, we propose an account in which core elements of the item-based theories hold and fixations are included.

ACKNOWLEDGMENT

The authors thank Dr. Michael A. Lebedev for his helpful comments on an earlier version of the manuscript.

Cognitive architecture enables comprehensive predictive models of visual search

doi:10.1017/S0140525X16000121, e142

David E. Kieras^a and Anthony Hornof^b

^aEECS Department, University of Michigan, Ann Arbor, MI 48109-2121;

^bDepartment of Computer and Information Science, University of Oregon, Eugene, OR 97403-1202.

kieras@umich.edu

<http://web.eecs.umich.edu/~kieras/>

hornof@cs.uoregon.edu

<http://ix.cs.uoregon.edu/~hornof/>

Abstract: With a simple demonstration model, Hulleman & Olivers (H&O) effectively argue that theories of visual search need an overhaul. We point to related literature in which visual search is modeled in even more detail through the use of computational cognitive architectures that incorporate fundamental perceptual, cognitive, and motor mechanisms; the result of such work thus far bolsters their arguments considerably.

Hulleman & Olivers (H&O) help move the study of visual search in the right direction by promoting the idea of *active vision* (Findlay & Gilchrist 2003), which emphasizes the role of eye movements in visual search over a traditional emphasis on covert shifts of attention; this also reflects a shift to eye movement measures rather than response times as the relevant dependent variables. Their paper incorrectly suggests, however, that a consideration of visual objects as discrete psychological objects is problematic. The problem is not that items are involved, but rather that most current theories assume a simplistic relationship between number of items and reaction time that misrepresents how visual capabilities are actually used in the task. Therefore, for many stimuli and tasks such as searching for icons on a computer screen (e.g., Fleetwood & Byrne 2006; Kieras & Hornof 2014), the goal is to identify a specific display object (e.g., by clicking on it) which implies that the search must be item-based at

some level, even though the subject might be processing multiple visual objects simultaneously.

We believe that the actual theoretical problem that underlies the issues pointed out by H&O is an impoverished theoretical framework for how human perceptual, cognitive, and motor mechanisms operate together in performing visual search. Although their model demonstrates an approach that improves on the status quo, the problem can be more thoroughly alleviated by the use of an explicit computational cognitive architecture foundation.

In an analogy with computer architecture, a computational cognitive architecture such as ACT-R (Anderson & Lebiere 1998) or EPIC (Meyer & Kieras 1997) proposes basic "hardware" mechanisms for perception, cognition, and action, with task-specific "software" in the form of production rules that provide a strategy for how to perform the task with the architectural mechanisms. In EPIC, the visual modules in fact already incorporate the *functional viewing field* (FVF) concept that H&O promote. These modules use functions from the psychophysical literature for the detectability of visual properties based on object eccentricity and size (e.g., Anstis 1974; Gordon & Abramov 1977; Virsu & Rovamo 1979). Certain visual search results characterize the memory stores for visual information (e.g., Henderson & Castelhano 2005; Kieras 2011; Peterson et al. 2001), and the oculomotor module incorporates results from eye movement studies of the speed and accuracy of saccades (e.g., Abrams et al. 1989; Harris 1995). The production rules comprising the cognitive strategy specify how the human uses the perceptual, motor, and memory resources to accomplish the task. For example, the strategy determines which object should be fixated next based on the task requirements and the available perceptual and memory information, and what is to be done if the fixation has failed, such as missing the targeted object. EPIC's cognitive strategy represents explicitly the cognitive decisions that must be executed to complete visual search tasks, the same sort of decisions that H&O summarize in the "flow diagram of the conceptual framework" shown in their Figure 2.

Unlike simple ad hoc models, a computational cognitive architecture serves to synthesize many empirical effects in a stable, reusable fashion in the architecture components, and because the strategy is represented separately, it is easy to build a model for a different task, or to explore the consequences of different strategies for the same task, all based on the same architectural assumptions. The resulting simulation models provide a computational version of the active vision concept itself and can be applied to practical problems such as analyzing and improving the design of computer interfaces and multimodal interaction (Halverson & Hornof 2011; Hornof 2004; Kieras & Hornof 2014; Kieras & Meyer 1997).

However, a lesson of cognitive architecture modeling is that even in simple tasks, subjects can adopt subtle strategies that can obscure the underlying mechanisms unless taken into account (Kieras & Meyer 2000; Meyer & Kieras 1999; Zhang & Hornof 2014). In our models of visual search, the task strategy plays a role in determining the object to be next fixated and, when there are no clear candidates for the next saccade, there are a variety of possible decisions that could be made, and these decisions affect the predicted search performance.

Unless some effort has been made to control subject strategies in the experiment, the data themselves may be an arbitrary mixture of idiosyncratic performance by each subject – and this strategy indeterminism can happen even in seemingly simple tasks (Meyer & Kieras 1999; Kieras & Meyer 2000; Schumacher et al. 2001). Modeling such data in the aggregate typically requires either ad hoc or unrepresentative strategies, but the alternative of identifying individual subject strategies in normal-size data sets is extremely difficult. Instead, providing explicit incentives to the subjects can lead to more stable strategy choices and thus data that reflect underlying mechanisms much more clearly (e.g., Thompson et al. 2015). Thus, we believe that task strategy

needs to be considered more explicitly both in experiments and models about visual search.

We applaud the authors for their effort to further promote gaze fixations and the FVF as first-class entities in models and theories of visual search. We believe that the development of computational cognitive architecture models provides a promising pathway to achieve the goals they have expressed. We encourage other researchers to embrace these positive developments but to also go further and (a) more explicitly consider the role of cognitive strategy in visual search and (b) as hinted by H&O, collect the empirical data needed to describe more completely and parametrically how visual properties are detected based on object eccentricity, size, and density (building on Anstis 1974; Bouma 1970; Engel 1977; Gordon & Abramov 1977; Virsu & Rovamo 1979). Both are needed for a comprehensive predictive model of visual search.

How functional are functional viewing fields?

doi:10.1017/S0140525X16000133, e143

Árni Kristjánsson, Andrey Chetverikov, and Manje Brinkhuis

Faculty of Psychology, School of Health Sciences, University of Iceland, 101, Reykjavík, Iceland.

ak@hi.is andrey@hi.is manjebrinkhuis@gmail.com

Abstract: Hulleman & Olivers' (H&O's) proposal is a refreshing addition to the visual search literature. Although we like their proposal that fixations, not individual items should be considered a fundamental unit in visual search, we point out some unresolved problems that their account will have to solve. Additionally, we consider predictions that can be made from the account, in particular in relation to priming of visual search, finding that the account generates interesting testable predictions.

Hulleman & Olivers' (H&O's) target article is a refreshing addition to the visual search literature. We agree with them that there is need for a more flexible conception of visual search, and that eye movements should not be considered a nuisance factor. They are, however, not the first to point out problems with what they call the *item-based* approach, where slopes of set-size and response times take center stage. Concerns about traditional visual search approaches are raised in parallel models of visual search (Eckstein 1998; Kristjánsson 2015; Palmer et al. 1993) showed how slopes are ambiguous measures of search behavior; and Wang et al. (2005) have shown how even very difficult searches can yield flat slopes, calling for changed conceptions of search. But as H&O rightly highlight, satisfactory replacements to traditional approaches have not surfaced.

Functional viewing fields (FVF) play a central role in their account. Although we think this approach is useful, we still feel it comes up short on some important questions. Perhaps against the authors' intention, FVFs may conveniently describe a continuum between easy search involving the whole visual field ("parallel," broad, shallow processing within saliency maps) and item-based processing ("serial," narrow but deeper), similar to an "attentional window" (Belopolsky et al. 2007), whose size scales with attentional load (Lavie et al. 2004). The "parallel" versus "serial" dichotomy may no longer be useful for developing new ideas (Kristjánsson 2015; Nakayama & Martini 2011). FVFs are spatially constrained, and so the concept may encounter similar problems as spotlight metaphors. Attending to multiple moving items (Cavanagh & Alvarez 2005), perceptual grouping (Kerzel et al. 2012; Vatterott & Vecera 2015), or predictability (Jefferies et al. 2014) can shape or divide the attentional window, arguing against the idea of a single FVF. Additionally, whether items within spatially constrained FVFs are processed in parallel is not clear. For example, priming studies demonstrate that attention spreads unevenly between

targets and distractors within FVFs (Kristjánsson & Driver 2008). A single FVF (even with a dynamically changing size) is therefore unlikely to explain nonuniform or spatially noncontiguous attention distribution.

Sometimes H&O seem to try and explain the *literature* on visual search rather than actual visual search and attention. One example is that FVFs may be difficult to define operationally, while they rather straightforwardly explain set-size effects. FVFs are supposedly small in difficult search tasks, but determining which tasks are hard seemingly requires set-size slopes, which FVF size is supposed to account for. This is circular. H&O discuss other factors influencing the size of FVF (e.g., distractor heterogeneity), but whether FVFs add to the explanatory power already provided by these factors is unclear. The proposal does, in other words, not contain a clear way of predicting FVF size except with already well-known tools.

According to H&O, set-size effects are explained with fixations, and they explicitly assume no covert attentional shifts within FVFs. Search where eye movements are *not* allowed should therefore not yield such effects when distractors are isoeccentric. But set-size effects persist when eccentricity is controlled for and eye movements are eliminated, (e.g. Carrasco et al. 2001; Foley & Schwarz 1998; Palmer et al. 1993). Rather, set-size effects might reflect the *discriminability* of target versus distractors, which relies on set-size, covert attention, and position within FVFs (Anton-Erxleben & Carrasco 2013; Carrasco et al. 2001; Carrasco & Yeshurun 1998). Importantly, if target location is pre-cued set size effects are reduced (Carrasco et al. 2001; Foley & Schwarz 1998), which neither item-based selection, nor FVF's can explain. We agree that target selection can rely on discriminability between items processed in parallel within FVFs, but the best approach to explaining how we attend in the visual scene will probably be multifaceted, involving covert and overt attentional shifts.

Despite these criticisms H&O's proposal is refreshing. We suggest several predictions that can be made from it. We consider priming of visual search (Maljkovic & Nakayama 1994; see Kristjánsson & Campana [2010] for review). Such priming occurs for searches of varying difficulty (Ásgeirsson & Kristjánsson 2011) and according to H&O, search difficulty determines FVF size. If stimuli are predominantly processed within FVFs, then for priming to manifest its effects, a primed target must fall within the FVF. Increased search difficulty contracts the FVF, lowering the probability that a target will fall within it. For difficult search tasks, priming effects should therefore decrease when set-size increases, while for easy tasks they should be constant (or decrease more slowly), as the FVF is larger and therefore likely to include the target. Here the proposal generates testable hypotheses, where the literature does not have clear answers (but see Becker & Ansorge 2013). Analogously, priming effects for targets should also last longer for easy search than for difficult search. With smaller FVFs more fixations are required to find the target. Hence, there will be more intervening fixations between the ones that include the target, most likely leading to faster decay. Temporal profiles of priming have been investigated (Kruijne et al. 2015; Martini 2010; Brascamp et al. 2011), but these studies do not provide a clear test of this prediction. Notably, it runs counter to a recent proposal that priming of conjunction search is longer lasting than feature priming (Kruijne & Meeter 2015).

Finally, we ask whether FVF size primes from previous trials, though as we discuss above, the measurement of the FVF size is problematic. While it is debatable that this is a prediction unique to FVFs, the approach clearly predicts priming for fixations rather than individual items. Fuggetta et al. (2009) found that search was faster when the physical size (and set-size) of a search array was constant than when it changed, but priming of items versus fixations has not directly been contrasted. We hope that these and other new predictions will help with assessing the usefulness of H&O's new approach.

Gaze-contingent manipulation of the FVF demonstrates the importance of fixation duration for explaining search behavior

doi:10.1017/S0140525X16000145, e144

Jochen Laubrock, Ralf Engbert, and Anke Cajar

Department of Psychology, University of Potsdam, 14476 Potsdam, Germany

jochen.laubrock@uni-potsdam.de ralf.engbert@uni-potsdam.de

anke.cajar@uni-potsdam.de

<http://mbd.uni-potsdam.de/EngbertLab/Welcome.html>

Abstract: Hulleman & Olivers' (H&O's) model introduces variation of the functional visual field (FVF) for explaining visual search behavior. Our research shows how the FVF can be studied using gaze-contingent displays and how FVF variation can be implemented in models of gaze control. Contrary to H&O, we believe that fixation duration is an important factor when modeling visual search behavior.

Hulleman & Olivers (H&O) criticize the visual search literature for having focused largely on the individual item as the primary unit of selection. As an alternative to this view, the authors propose that (1) visual sampling during fixations is a critical process in visual search, and that (2) factors in addition to items determine the selection of upcoming fixation locations. H&O developed a very parsimonious simulation model, in which the size of the functional visual field (FVF) adapts to search difficulty. Items within the FVF are processed in parallel. Consequently, when search difficulty is very high, the FVF shrinks to a size of one item, effectively producing serial search. When search difficulty is lower, more items are processed in parallel within the FVF. These modeling assumptions were sufficient to qualitatively reproduce much of the canonical data pattern obtained in visual search tasks.

We applaud H&O for acknowledging the important and long-neglected contribution of eye movement control in guiding the search process, because we believe that many attentional phenomena can be explained by considering oculomotor activity (e.g., Laubrock et al. 2005; 2008). Although not all attention shifts are overt, the neural underpinnings of covert attention shifts are largely identical to those of eye movement control (Corbetta et al. 1998). Attention research should therefore be able to profit from the advanced models of the spatiotemporal evolution of activations in visual and oculomotor maps as well as from the methods for directly manipulating the FVF.

Gaze-contingent displays are a method to directly manipulate the FVF. For example, in the moving-window technique (McConkie & Rayner 1975) information is only visible within a window of variable size that moves in real-time with the viewer's gaze. Visual information outside of the window is either completely masked or attenuated. A very robust result from studies using this technique is that FVF size is globally adjusted to processing difficulty. In reading research, the size of the FVF is often called the *perceptual span*, which has been shown to increase with reading development (Rayner 1986; Sperlich et al. 2015) and to be dynamically adjusted, for example, when viewing difficult words (Henderson & Ferreira 1990; Schach & Engbert 2012). In scene perception parametrically increasing peripheral processing difficulty, by, for example, selectively removing parts of the spatial frequency spectrum from the peripheral visual field (Fig. 1, top), leads to corresponding reductions in saccade amplitudes (Cajar et al. 2016a; Loschky & McConkie 2002), suggesting a smaller FVF. These modulations are stronger when broad features are removed than when fine details are removed (Cajar et al. 2016a), reflecting the low spatial resolution of peripheral vision. Conversely, when the filter is applied to the central visual field (Fig. 1, bottom) saccade amplitudes increase—particularly if fine detail is removed, corresponding to the high spatial resolution of foveal vision. Cajar et al. (2016b) show that these very robust modulations of mean saccade



Figure 1 (Laubrock et al.). Illustration of gaze-contingent spatial frequency filtering in real-world scenes. The white cross indicates the current gaze position of the viewer. Top: Peripheral low-pass filtering attenuates high spatial frequencies (i.e., fine-grained information) in the peripheral visual field. Bottom: Central high-pass filtering attenuates low spatial frequencies (i.e., coarse-grained information) in the central visual field.

amplitude are directly correlated with the distribution of attention (i.e., the perceptual span).

Are existing models of saccadic selection compatible with a variable FVF? In biologically plausible models, a critical feature is a spatial map with a limited spotlight of attention (i.e., an FVF-like representation). Additionally, a simple memory mechanism (called *inhibitory tagging*) prevents the model from getting stuck by continually selecting the point of highest saliency. Engbert and colleagues implemented such a dynamic model of eye guidance in scene viewing (Engbert et al. 2015), based on an earlier model of fixational eye movements (Engbert et al. 2011). The combination of two interacting attentional and inhibitory maps could reproduce a broad range of spatial statistics in scene viewing. Whereas these models do explain the selection of fixation locations fairly well, an additional mechanism that adjusts the zoom lens of attention with respect to foveal processing difficulty (Schach & Engbert 2012) is necessary to capture modulations of fixation duration.

In comparison to the complexity of these detailed dynamic models, the H&O model has the advantage of simplicity. However, this comes at a cost of somewhat unrealistic assumptions. For example, H&O assume that fixations have a constant

duration of 250 ms and that only the number and distribution of fixations adapt to search difficulty. The authors justify this decision with previous research that barely found effects of target discriminability on fixation durations in typical search displays. However, at least for visual search in complex real-world scenes, research shows that fixation durations are indeed affected by search difficulty (e.g., Malcolm & Henderson 2009; 2010).

Thus, not only selection of fixation locations, but also control of fixation duration is influenced by the FVF. In particular, mean fixation duration increases when visual information accumulation in regions of the visual field is artificially impaired by means of gaze-contingent spatial filtering (Laubrock et al. 2013; Loschky et al. 2005; Nuthmann 2014; Shioiri & Ikeda 1989). However, this effect is observed only when filtering does not completely remove useful information—otherwise, default timing takes over, meaning that fixation durations fall back to the level observed during unfiltered viewing (e.g., Laubrock et al. 2013). This might explain why effects of visual search difficulty are more often reported for number of fixations rather than fixation duration. A critical aspect of a model of fixation duration in visual scenes is parallel and partially independent processing of foveal and peripheral information (Laubrock et al. 2013). Given that both FVF size and fixation duration adapt to task difficulty, an important research goal of the future is to integrate models of fixation location and fixation duration.

Set size slope still does not distinguish parallel from serial search

doi:10.1017/S0140525X16000157, e145

Daniel R. Little,^a Ami Eidels,^b Joseph W. Hout,^c and Cheng-Ta Yang^d

^aMelbourne School of Psychological Sciences, The University of Melbourne, Parkville VIC 3010 Australia; ^bSchool of Psychology, The University of Newcastle, Callaghan NSW 2308; ^cDepartment of Psychology, Wright State University, Dayton, OH 45435-0001; ^dDepartment of Psychology, National Cheng Kung University, Tainan City 701, Taiwan (R.O.C.).

daniel.little@unimelb.edu.au

<http://www.psych.unimelb.edu.au/people/daniel-little>

ami.eidels@newcastle.edu.au

<http://www.newcl.org/eidels>

joseph.hout@wright.edu

<http://www.wright.edu/~joseph.hout/>

yangct@mail.ncku.edu.tw

<http://vcmlab.psychology.ncku.edu.tw/vcmlab/>

Abstract: Much of the evidence for theories in visual search (including Hulleman & Olivers' [H&O's]) comes from inferences made using changes in mean RT as a function of the number of items in a display. We have known for more than 40 years that these inferences are based on flawed reasoning and obscured by model mimicry. Here we describe a method that avoids these problems.

In their recent review, Algom and colleagues note that “Generations of cognitive psychologists appear to have been rendered oblivious to the developments in mathematical psychology on the importance and (im)possibility of distinguishing between parallel and serial processing based on straight line mean RT functions” (Algom et al. 2015, p. 88). Although H&O make a number of cogent points regarding the importance of eye movements, the nature of eccentricity and target salience, and the role of functional viewing field (FVF), when it comes to their discussion of serial and parallel search, the authors unfortunately repeat the error of overinterpreting mean RT set-size functions.

As far back as Townsend (1971), researchers have demonstrated that a parallel search process can lead to increases in RT as a function of set size, whereas serial search can lead to flat RT slopes as a function of set size. Consequently, the

authors' attack on item-based search, which in the visual search literature is synonymous with serial search, uses the set size-RT slope to make erroneous inferences about processing. This is an ongoing issue with visual-search data, the ambiguity of which was so great that Wolfe (1998b) declared the serial/parallel distinction to be a dead end and initiated a switch in terminology calling zero-slope functions “efficient” and positive-slope functions “inefficient” search, effectively masking but not resolving the problem.

Townsend (1972; see also Townsend & Ashby 1983) pointed out that a parallel model can mimic a serial model by setting the intercompletion times of items in a parallel race (i.e., the unobservable time that a parallel racing item completes its race) equal to finishing times of items in a serial process. The necessary implication is that relying on mean RTs as a function of set size (in any search task, visual or memory) does not have the inferential or discriminatory ability to differentiate serial from parallel processing. As noted by Townsend (1990) and acknowledged by Wolfe et al. (2010b), by utilizing factorial designs and estimating RT distributions, serial and parallel models (and several other important classes of processing models) can be distinguished.

This theory and method, collectively known as Systems Factorial Technology (SFT), was fully developed and introduced 20 years ago by Townsend and Nozawa (1995), and it continues to be applied, developed, refined, and extended (Eidels et al. 2011; Hout & Townsend 2012; Little et al. 2015; Townsend & Wenger 2004). SFT can differentiate serial and parallel processing by analyzing RT distributions from conditions that vary the strength or quality of the stimulus to slow down or speed up processing along each of the two dimensions (e.g., signal-modality—audition and vision, or signal location—top and bottom). Crossing two factors with two levels of strength, we obtain four conditions: LL (low salience on both dimensions), LH and HL (low salience on one dimension and high salience on the other), and HH (high salience on both dimensions). Diagnostic contrasts are computed by combining the RT distributions (i.e., survivor functions) from the four factorial conditions. Each architecture (e.g., serial or parallel) makes different predictions for the diagnostic contrasts. Serial models predict additivity; that is, the change from LL to HH should equal the sum of the changes on each dimension separately; hence, $(LL-LH)-(HL-HH) = 0$. By contrast, parallel models predict overadditivity (i.e., positive, for self-terminating processing) or underadditivity (for exhaustive processing). Inhibitory and facilitatory models also predict under- and overadditivity, respectively (Eidels et al. 2011). These nonparametric tests allow for entire classes of models to be tested and falsified. For instance, a completely negative contrast rules out all serial models.

SFT also provides the ability to differentiate many other important facets of information processing in addition to architecture, including workload capacity (how processing efficiency changes in response to changes in the number of targets to process), (in)dependence (whether processing channels are mutually facilitatory or inhibitory), and stopping rule (whether processing is self-terminating or exhaustive). The last property is of particular importance to the authors' simulation because many of their results depend on the stopping rule.

These methods are particularly useful in verifying aspects of computational theories such as the one proposed by the authors. The authors assume that items within the FVF are processed in parallel and that the size of the FVF can be inferred by examining the slope of RT set-size functions. Like the theories the authors are attempting to displace, this procedure again requires too much of the set-size function. The focus on RT variability is more promising, and we applaud the general approach of breaking down the search tasks by difficulty and examining target-present and target-absent variability. However, the assumption of deterministic fixation duration influences the conclusions that the authors draw from variability. Even assuming a variance in fixation duration that is independent of task difficulty and FVF size, the

model will predict higher variance with more fixations. That could mitigate, if not wash out, the decrease in variability with the highest difficulty. Ideally one would attempt to directly manipulate target-present and target-absent RT variability, which applications of SFT to visual search have done (see e.g., Fific et al. 2008; Sung 2008).

Could SFT be used to examine the properties of the FVF? Yes, and easily. One would only need to manipulate the detection difficulty of two targets in an array that either did or did not require eye movements. Related work by one of the authors (C-TY) using SFT in redundant target detection has shown that processing is parallel self-terminating, and of limited capacity when there are few eye movements. By contrast, when eye-movements to a target are forced, processing instead conforms to serial processing (at least for some observers; see also Fific et al. 2010). We believe that these results lend preliminary support to the authors' inferences but rely on the more rigorous methods of SFT rather than the perilous mean RT set-size function slopes.

Oh, the number of things you will process (in parallel)!

doi:10.1017/S0140525X16000169, e146

Alejandro Lleras, Deborah A. Cronin, Anna M. Madison, Marshall Wang, and Simona Buetti

University of Illinois at Urbana-Champaign, Champaign, IL 61820.

alejandrolleras@gmail.com cronin2@illinois.edu
amadiso2@illinois.edu zhiyuanwang42@gmail.com sbuetti@gmail.com
<http://www.psychology.illinois.edu/people/alleras>

Abstract: We highlight the importance of considering the variance produced during the parallel processing stage in vision and present a case for why it is useful to consider the “item” as a meaningful unit of study when investigating early visual processing in visual search tasks.

There is much to like about Hulleman & Olivers' (H&O's) proposal. However, the article falls short on at least two fronts. Mostly, it suffers from over-generalizations in its core assumptions that limit the potential of the article.

First, it assumes that early visual computations are identical, irrespective of the type of visual search an observer is asked to perform. However, there is strong evidence that the type of early computations performed by the visual system are fundamentally different when searching for a known target (e.g., look for a “T”) than when looking for an unknown target (as in oddball search), even when that unknown target “pops-out” from the distractors by virtue of its features (e.g., Bravo & Nakayama 1992; Buetti et al. 2016; Lamy & Kristjánsson 2013; also see Li et al. 2004; 2006, for neural evidence that top-down goal changes early visual processing).

Second, ample evidence exists that *much* can be performed in searches that do not require eye movements. The authors acknowledge, then quickly dismiss, this observation by assuming all “easy” search (i.e., parallel search) can be accounted for by simply assuming a very large FVF. On this front, H&O's proposal is no better than previous proposals that assume all parallel searches are created equal (e.g., Wolfe 1994). They are not. What is remarkable is that at that scale – that is, in visual searches that are performed in parallel and without the need for eye movements – the “single item” is a meaningful unit of measurement: For a fixed-target search, RTs increase logarithmically as a function of the number of items and the steepness of that logarithmic curve is determined by the similarity between the target and the individual items (Buetti et al. 2016). The result of glossing over the subtleties of parallel search is that H&O's remains very much a univariate approach to visual search: determining

the FVF (or the size of the pooling region, as in Rosenholtz' work) should be all that is needed to understand search performance in any situation. Dismissing very efficient searches as not interesting seems to us to miss an important point. In the real world, peripheral vision *can* and probably *does* make very fast and accurate decisions about many regions/pooling regions/textures/items/objects because it has sufficient information to determine which ones are unlikely to contain the target of the search (Balas et al. 2009; Rosenholtz et al. 2012b, though see other work challenging the notion of peripheral “pooling” or averaging regions, Ester et al. 2013; 2015). Our work shows that these peripheral decisions come at an detectable performance cost. That is, in addition to the serial processing mechanism imposed by successive moves of the FVF proposed by the authors, an additional source of variance in performance determines the time it takes to reach decisions about individual peripheral items within the parallel processing stage. Visual search is (at the very least) a bi-variate problem: one source of variance determines the number of serial steps in processing (the authors propose the size of the FVF), and another source determines the efficiency with which individual items are judged in the parallel process. This was highlighted in four experiments in Buetti et al. (2016): When search displays contain both easy-to-reject items (*lures*) and need-to-scrutinize items (*candidates*), one can isolate the logarithmic contribution to RTs that arise from parallel processing (i.e., the rejection of lures) from the linear contribution to RTs induced by serial processing (i.e., the scrutiny of candidates, or as the authors propose, the number of moves of FVF). Figure 1 illustrates how both sources of variances can be disentangled and visualized by plotting separate RT functions for conditions containing an identical number of candidates.

Finally, it is quite unlikely that fixations are random, as proposed by the authors. They are likely determined by the output of the computations in areas outside of the FVF, as proposed by models such as Zelinsky's TAM (2008), for example, and performed mostly by parallel processing as well.

In sum, though we agree with the sentiment that overly-focusing on the “single-item” has perhaps lead astray researchers interested in inefficient search, we anticipate a revival of interest on the single item as meaningful for understanding search behavior. This revival will come not where most would have expected (or where most have looked) – in serial/slow searches – but rather precisely where most (including H&O) have ignored: in parallel search. This follows because in the context of parallel visual search, manipulating the number of (high signal-to-noise ratio) items in the periphery allows for a precise quantification of the efficiency of parallel processing and of the similarity between the peripheral items and the search template. Of course, one might wonder whether this is at all relevant to our understanding of real-world visual search. Given the visual heterogeneity of a real world scene, the number of items that *ought* to be closely inspected by focused attention is likely to be only a fraction of the total (Neider & Zelinsky 2008; Wolfe et al. 2011a). Take the simple example of looking for lawn furniture in your garden: in spite of there being an very large number of items in the scene (flowers, trees, grass, animals, etc.), most of them are vastly different from lawn furniture and one would never spend time closely attending to them when looking for a place to sit. Yet, as our research has shown, the presence and visual attributes of these not-to-be-inspected items *do* affect the time it will take observers to find a place to sit.

Nonetheless, these shortcomings are clearly fixable and a better account of the contribution of parallel vision to behavioral performance can be easily integrated into the H&O proposal. Future empirical work should be aimed at estimating the contribution of peripheral processing both outside of FVF and within FVF to (a) planning future eye movements and (b) predicting fixation processing times.

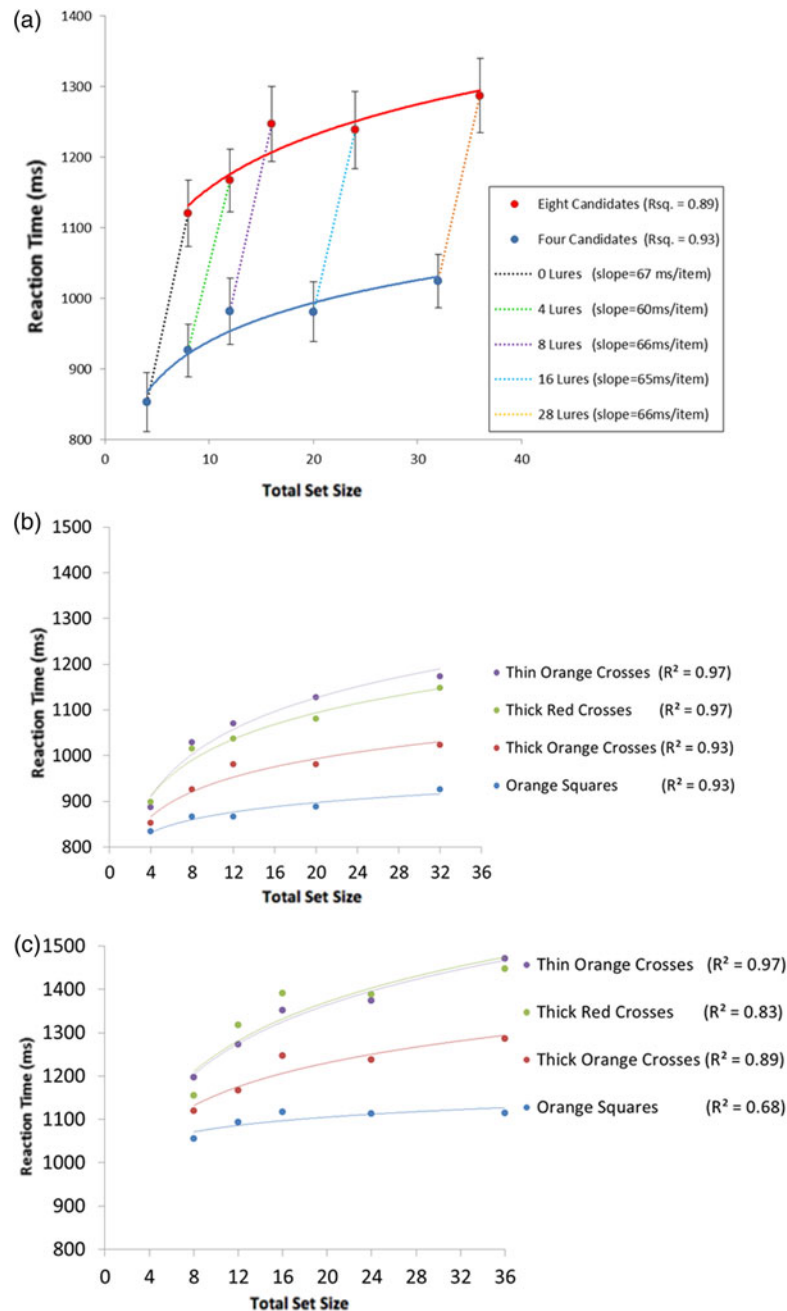


Figure 1 (Lleras et al.). Results from Experiments 3A–D in Buetti et al. (2016) showing time (in ms) to find a target (an oriented red T) as a function of the number of elements in the display, shown separately for displays containing 4 and 8 candidates (oriented red Ts), amongst a varying number of lures. The full lines show the best fitting logarithmic trend for each series, and the corresponding measure of fit (R^2). **A.** Data from Experiment 3A: The dotted lines visualize the scrutiny functions for each level number of lures. The slopes for the scrutiny function when no lures were presented (0 lures) was 67 ms/item and was no different than when 4, 8, 16, or 28 lures were present. Error bars indicate the between-subject standard error of the means. **B–C.** Combined data from Experiments 3A–D showing logarithmic screening functions when 4 (B) or 8 (C) candidates are present in the display, showing orderly logarithmic sensitivity to target–lure dissimilarity.

The FVF framework and target prevalence effects

doi:10.1017/S0140525X16000170, e147

Tamaryn Menneer, Hayward J. Godwin, Simon P. Liversedge, Anne P. Hillstrom, Valerie Benson, Erik D. Reichle, and Nick Donnelly

University of Southampton, Southampton, Hampshire, SO171BJ, United Kingdom.

T.Menneer@soton.ac.uk hayward.godwin@soton.ac.uk
S.P.Liversedge@soton.ac.uk anne.hillstrom@gmail.com
V.Benson@soton.ac.uk E.Reichle@soton.ac.uk
N.Donnelly@soton.ac.uk

Abstract: The Functional Visual Field (FVF) offers explanatory power. To us, it relates to existing literature on the flexibility of attentional focus in visual search and reading (Eriksen & St. James 1986; McConkie & Rayner 1975). The target article promotes reflection on existing findings. Here we consider the FVF as a mechanism in the Prevalence Effect (PE) in visual search.

Hulleman & Olivers' (H&O's) focus on simulating how slope gradients are influenced by the difficulty of search provokes a thoughtful discussion. However, limiting simulations to these data alone can mistakenly suggest that the FVT framework's usefulness is itself limited. To help address this perceived limitation, we have reanalysed data from a study of the Prevalence Effect (PE; Godwin et al. 2015a; Wolfe et al. 2005).

The PE refers to the influence that target probability has on both target selection and verification (e.g., Godwin et al. 2015b; Hout et al. 2015). Frequently occurring targets tend to be found and verified quickly. In contrast, their absence is reported slowly. The presence of infrequent targets is reported slowly and their absence reported quickly.

The target article accounts for the modulating effect of target discriminability on search reaction times solely by changes in the size of the FVF. Might changes in the size of the FVF also contribute to the PE? Specifically, high target prevalence might lead participants to initially adopt a broader FVF than when target prevalence is low. A relatively broad FVF would allow the presence of targets to be detected quickly whereas a relatively narrow FVF would lead to slowed target detection. In deriving these hypotheses, we have made two assumptions. First, and to account for slow target-absent responses when target prevalence is high, we assume that failure to find evidence of target presence when the FVF is broad leads to a dynamic resizing of the FVF to allow, at the limit, item-by-item analysis (note that a global-to-local fixation pattern is consistent with recent consideration of search, Godwin et al. 2014; Over et al. 2007). Second, we assume that the fixation point of a broadened FVF is more likely to be centrally than peripherally positioned. For a broad FVF, a central fixation will encompass more items than a non-central fixation will. These reduce to a hypothesis that, early in search, fixations are more centrally biased in high-prevalence than low-prevalence search.

To test this hypothesis, we reanalysed data on target-present trials from Godwin et al. (2015a). Space restrictions prohibit a full account of these data and analyses. Briefly, to assess the patterns quantitatively, the distribution of fixation locations across displays were normalised within high- and low-prevalence conditions and split into fixations made early and late in search (as defined by median split). Z-scores for the differences between high- and low-prevalence conditions were calculated for these normalised data. We found that increasing prevalence is associated with more fixations to the centre of search displays early in search. A centre bias (Tseng et al. 2009) is present early in low-prevalence search, but the bias is significantly stronger under high prevalence.

These data, then, are consistent with the FVF framework. However, we do not claim that our reanalysis provides unequivocal support. Rather, the framework prompts us to reconsider data in a way that provide an additional account of how search patterns might change with target prevalence.

The current utility of the FVF is tempered, in our view, by two limitations. First, the current focus on numbers of fixations ignores the influence of fixation duration. Increasing cognitive demands affects both the number and duration of fixations (Liver- sedge & Findlay 2000). Consequently, any comprehensive framework of search behaviour must explain both fixation number and duration. Recent evidence suggests that fixation durations during visual search are controlled on the basis of a trade-off between making rapid fixations and allowing time to examine objects in the display (Godwin et al. in press). As a consequence, there have been calls for a greater understanding of fixation duration variability during visual search tasks (Reingold & Glaholt 2014).

Second, the authors rightly wish to extend consideration to searching in scenes. As the search environment becomes richer in contextual information, equation of selection time, processing time, and dwell time to fixation time (sect. 6.3) becomes more challenging. In reading, *spillover effects* are frequently observed (whereby a linguistic influence of one word is seen to affect fixations on it and later words in the sentence; Rayner & Duffy

1986). By extension, visual search in scenes may also be subject to partial dissociation between fixation location and the set of locations from which information is currently being processed. To this extent, evaluation of effects across temporally contiguous fixations as well as spatially contiguous fixations is a critical issue for theoretical development.

In sum, we consider the FVF framework as a useful prompt to rethink visual search. Here we have provided some provisional data that might further support this framework. In addition, two areas of concern to be addressed in future developments have been noted.

An appeal against the item's death sentence: Accounting for diagnostic data patterns with an item-based model of visual search

doi:10.1017/S0140525X16000182, e148

Rani Moran,^{a,b,c} Heinrich René Liesefeld,^d Marius Usher,^c and Hermann J. Müller^{d,e}

^aMax Planck UCL Centre for Computational Psychiatry and Ageing Research, London WC1B 5EH, United Kingdom; ^bWellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom; ^cSchool of Psychological Sciences and Sagol School of Neuroscience, Tel Aviv University, Ramat Aviv, Tel-Aviv 69978, Israel; ^dDepartment Psychologie, Ludwig-Maximilians-Universität München, D-80802 Munich, Germany; ^eDepartment of Psychological Sciences, Birkbeck College, University of London, London WC1E 7HX, United Kingdom.

rani.moran@gmail.com

<https://iris.ucl.ac.uk/iris/browse/profile?upi=RMORA40>

heinrich.liesefeld@psy.lmu.de

http://www.psy.lmu.de/exp/people/ma/liesefeld_hr/

marius@post.tau.ac.il

<https://en-social-sciences.tau.ac.il/profile/marius>

hmueller@psy.lmu.de

<http://www.psy.lmu.de/exp/people/prof/mueller/>

Abstract: We show that our item-based model, competitive guided search, accounts for the empirical patterns that Hulleman & Olivers (H&O) invoke against item-based models, and we highlight recently reported diagnostic data that challenge their approach. We advise against “forsaking the item” unless and until a full fixation-based model is shown to be superior to extant item-based models.

Hulleman & Olivers (H&O) propose that fixations, rather than items, should serve as the central unit of analysis in theories of visual search. We agree that the fixation-based approach highlights important issues, especially the potential to integrate theories of manual responses and eye fixations. However, we disagree with H&O that their simulation “provides a compelling argument for abandoning” (sect. 6, para. 2) the item-based stance.

First, H&O emphasize the “difficulties or even the failure of these models to capture the distributional aspects of RTs” (sect. 6.1, para. 2). Recently we developed an (item-based) serial search model, dubbed *competitive guided search* (CGS; Moran et al. 2013), which successfully accounts for benchmark RT distributions (Wolfe et al. 2010b) and error rates across three classical search tasks (feature, conjunction, and spatial configuration [“2 vs. 5”] searches) and which is superior to a more flexible parallel model (Moran et al. 2016). Comparing our Figure 1 with H&O's Figures 3 and 4, we see that CGS can account remarkably well for all of the empirical patterns (see our Table 1 for simulation parameters). In fact, whereas H&O's simulation grossly misestimates some aspects of the data (it underestimates RTs in the easy and medium tasks and overestimates the target absent [TA] slope in the hard task and the rate of increase in SD in all tasks), CGS provides accurate predictions.

H&O especially highlight “variance inversion” (RT is more variable in TA for easy and medium-difficulty tasks but more variable in target-present [TP] displays for hard tasks) as evidence against

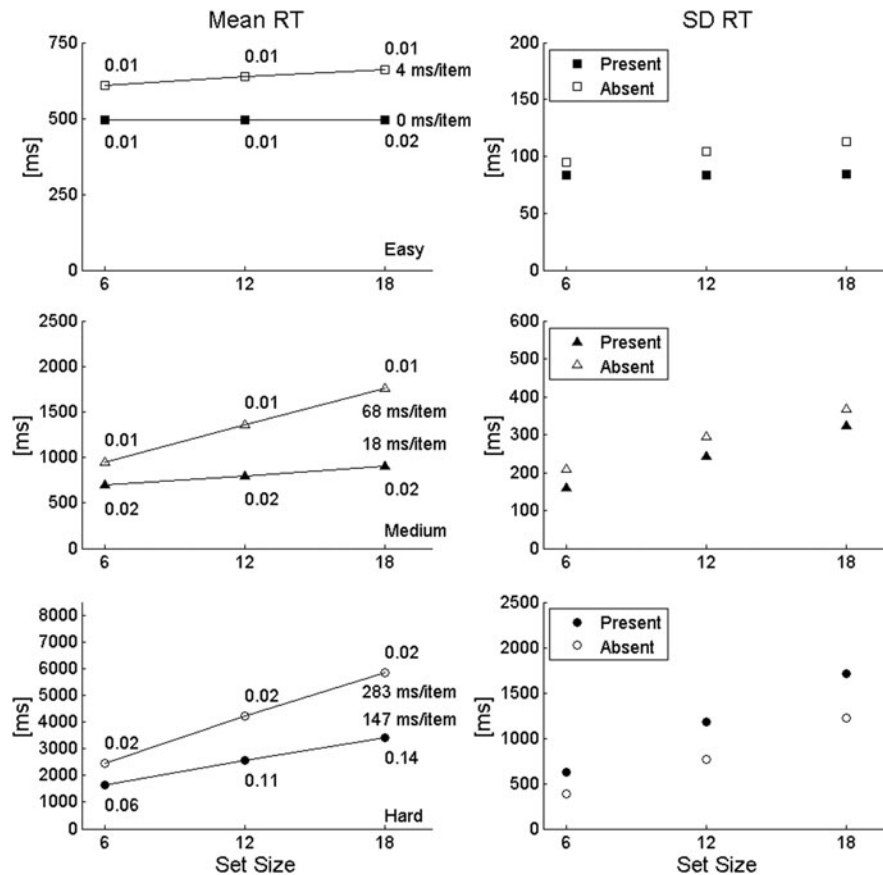


Figure 1 (Moran et al.). Correct RTs from three simulations of the CGS model (Table 1), reflecting hard, medium, and easy search tasks. Left and right panels depict means and SDs, respectively, as functions of set size. In the left panels, error proportions (next to the symbols) and search slopes (on the right) are indicated.

item-based models. However, CGS can explain this pattern in that multiple sources contribute to RT variability: (a) variance in how many items are identified before a response is issued (as determined by the tendency to quit the search [for TA] and by the amount of guidance towards the target [for TP]), and (b) variance in item identification time. Because (b) builds up as more items are identified and because, in all three simulations, more items are identified in TA than in TP, (b) contributes to a higher TA variability. However, in the “hard” task, (a) is so much larger for TP that it overrules the influence of (b) and, hence, inverses the variance. Finally, H&O’s claim that item-based models encounter the problem that slopes are much lower than expected based on estimates of attentional dwell time from other paradigms (~200–300 ms). A fourth simulation (Table 1, bottom row) with an attentional dwell time of 200 ms (simulated by identification time per item; Table 1, rightmost column) yielded slopes in the range claimed to be problematic for item-based models (25 and 73 ms/item for TP and TA, respectively). These moderate slopes are obtained because adding 1 item to displays increases the mean number of identified items by only ~0.13 (TA) and ~0.36 (TP).

Second, manipulating target discriminability parametrically (via orientation contrast), we recently found several diagnostic data patterns, which we believe successful search theories should explain (Liesefeld et al. 2016): (a) an intermediate difficulty range (between medium and easy search), where search is efficient (e.g., 1 ms/item) for TP but inefficient (e.g., 15 ms/item) for TA, yielding TA/TP slope ratios much larger than 2; and (b) strong effects of discriminability on search intercepts (decreases >100 ms) in the efficient range. CGS accounts for (a) by large guidance (so the target is always selected first) and a low quit parameters (so the number of inspected items in TA displays increases with set size); and CGS accounts for (b) by a speed-up of item identification. However, these patterns raise challenges to H&O’s approach. Indeed, repeating their simulation over a wide range of the maximal Functional Viewing Field (FVF) where search is efficient for TP (<5 ms/item), we found (Fig. 2) that the simulated TA/TP slope ratio is always smaller than 2 (left panel). Furthermore, when search is efficient for both TA and TP displays, the simulated intercepts hardly change (only about 5 ms; right panel).

Table 1 (Moran et al.). CGS parameters used in the simulations (see Moran et al. 2013).

Task	w_{target}	δ	θ	Δw_{quit}	T_{min}^{yes}	T_{min}^{no}	γ	m	θ/δ
Easy	400	50	3	6	0.35	0.43	12	0.015	0.06
Medium	2.9	0.317	0.022	0.003	0.5	0.5	30	0.014	0.07
Hard	1.06	0.714	0.246	0.02	0.3	0.3	5	0.024	0.344
200-ms Dwell Time	5.5	1.222	0.244	0.25	0.2	0.2	30	0.014	0.2

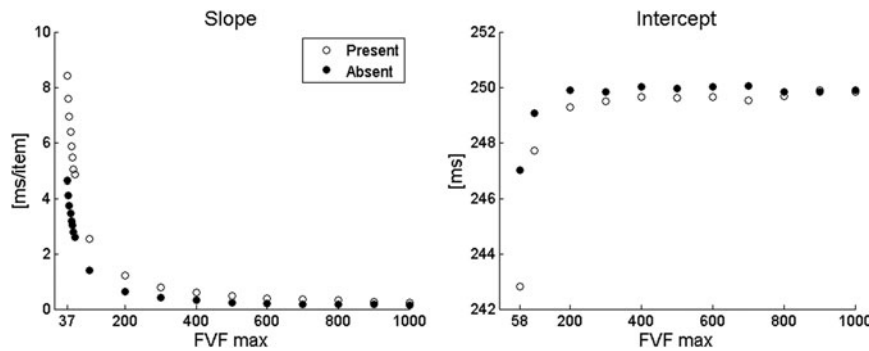


Figure 2 (Moran et al.). Simulated search slopes and intercepts as a function of the maximal Functional Viewing Field (FVF max) plotted in the range where TP (left) or both TA and TP (right) searches are efficient.

Finally, H&O concede that their high-level conceptual simulation provides merely a proof of principle for the viability of the fixation approach, but that there are still details that need to be explicated in a full model. It is tempting to think that the success of a model depends solely on its core functional assumptions and that, therefore, a fully explicated model would account for data better than the current preliminary framework. Alas, a model's success also hinges largely on peripheral assumptions and on their interaction with the central assumptions (e.g., Jones & Dzhafarov 2014). For example, H&O acknowledge a problem with their stopping rule. This rule is indeed peripheral to the focal items versus fixations debate; however, a stopping rule affects search-RT distributions and error rates substantially and it is therefore unclear how well H&O's framework will perform with a more plausible stopping rule.

In conclusion, we believe that the proposed framework would greatly benefit from developing the details of a full fixation-based model, followed by tests of how well it captures diagnostic empirical data patterns as compared to item-based models (using formal model comparisons). Unless and until this is done, however, we find reports of an “impending demise of the item” somewhat exaggerated.

Parallel attentive processing and pre-attentive guidance

doi:10.1017/S0140525X16000194, e149

Hermann J. Müller,^{a,b} Heinrich René Liesefeld,^a
Rani Moran,^{c,d,e} and Marius Usher^e

^aDepartment of Psychology, Ludwig-Maximilians-Universität München, Munich D-80802, Germany; ^bDepartment of Psychological Sciences, Birkbeck College, University of London, London WC1E 7HX, United Kingdom; ^cMax Planck UCL Centre for Computational Psychiatry and Ageing Research, London WC1B 5EH, United Kingdom; ^dWellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom; ^eSchool of Psychological Sciences and Sagol School of Neuroscience, Tel Aviv University, Ramat Aviv, Tel-Aviv 69978, Israel.

hmueller@psy.lmu.de

<http://www.psy.lmu.de/exp/people/prof/mueller/>

<http://www.bbk.ac.uk/psychology/our-staff/academic/hermann-muller>

heinrich.liesefeld@psy.lmu.de

http://www.psy.lmu.de/exp/people/ma/liesefeld_hr/

rani.moran@gmail.com

<https://iris.ucl.ac.uk/iris/browse/profile?upi=RMORA40>

marius@post.tau.ac.il

<https://en-social-sciences.tau.ac.il/profile/marius>

Abstract: This commentary focuses on two related, open questions in Hulleman & Olivers' (H&O's) proposal: (1) the nature of the parallel attentive process that determines target presence within, and thus presumably the size of, the functional visual field, and (2) how the pre-attentive guidance mechanism must be conceived to also account for search performance in tasks that afford no reliable target-based guidance.

Hulleman & Olivers (H&O) make an interesting case for an approach that takes eye fixations, rather than individual items, as its central unit. Within the fixational “functional field of view” (FFV), items are processed in parallel. The size of the FFV is adjusted according to search (target discrimination) difficulty, determining the number of fixations and thus RTs. While H&O's, and previous (e.g., Zelinsky 2008), arguments that eye movements and the FFV play a role in realistic visual search are persuasive, their model leaves (1) the *attentional* process that detects targets and (2) the pre-attentive process that guides fixations underspecified. Here, we discuss point (1) in relation to Humphreys and Müller's (1993) “Search via Recursive Rejection” (SERR) model (discussed by H&O in sect. 3.2), which, arguably, anticipated some of the ideas advocated by H&O, and (2) the need for a *pre-attentive* search-guidance mechanism in both SERR and H&O's model.

1. Like H&O's model, SERR deploys a sequence of parallel search steps to decide whether a target is present in the display. Although H&O are silent about the process that determines whether the target is present in each FFV region (a process their model considers as error-free), SERR—a connectionist implementation of Duncan and Humphreys' (1989) “Similarity Theory”—posits an error-prone mechanism. In SERR, items, the target and the distractors, within some FFV of spatially parallel processing compete for activating their (higher-level) template representations. When there are multiple distractors of the same complex feature description in the FFV, they are likely to win the competition over the single target, whereupon they are top-down suppressed “as a group.” This process operates recursively until either (1) the target activates its template, triggering a target-present (TP) decision; or (2) all items are “removed” from the FFV, leading to a target-absent (TA) decision. These dynamics are influenced by target–distractor similarity: The more similar the target is to (some of) the distractors, the more likely it is to be rejected along with a distractor group, yielding increasing miss rates. To bring the rate of target misses down to acceptable levels (matching those exhibited by humans), SERR must make several rechecking “runs” at the items in the FFV, until the target is either detected or consistently not found. Importantly, SERR produces miss rates that accelerate positively with the number of items in the FFV (especially with multiple distractor groups), in which case the rechecking strategy can become prohibitively expensive. As discussed by Humphreys & Müller (1993, p. 105), “A solution is to limit SERR's functional field so that there is a balance between the first-pass miss rate and the time cost incurred by rechecking”—providing an explicit, error-based “rule” for the FFV size adjustment. The adjusted FFV would then have to be deployed serially across the display (whether this involves covert or overt attention shifts). This resembles some of H&O's central ideas concerning discriminability-dependent FFV adjustments, which would be reflected in the number of attention shifts necessary to perform the task. As an aside, H&O are not quite right in stating that “the...empirical

work [associated with SERR] focused on relatively shallow search slopes" (sect. 3.2, para. 3): Müller et al. (1994) present simulations of human slopes (with slope estimates derived from simulated mean RTs and RT distributions) ranging, for example, in their Experiment 1, from about 30 to well over 200 ms/item.

2. Given a need for overt or covert attention shifts, efficient search would require an element of pre-attentive "guidance" for the FFV to be directed to (only) the most "promising" regions of the display. In principle, guidance can be provided by a combination of bottom-up and top-down mechanisms, for example, through the computation of local feature-contrast signals and their summation, across dimensions, on some search-guiding "overall-saliency" or "priority" map of the field. Note that this map is generally conceived as a pre-attentive representation, even though it is subject to top-down (feature- and dimension- as well as memory-based) biasing. Notions of guidance are at the heart of models from the Guided-Search (GS) family, including our "Competitive GS" model (e.g., Liesefeld et al. 2016; Moran et al. 2013; 2016), and well supported empirically. Although feature contrast computations themselves are not necessarily "item-based" (see, e.g., Itti & Koch 2001), much of what is known about their workings stems from item-based search experiments! Arguably, then, as acknowledged by H&O (in sect. 6.6), their model (and SERR!) would need to incorporate some notion of "guidance" to fully account for human search performance – which would bring it closer into line with "traditional," two-stage models of visual search like GS.

Note that H&O "buy in" guidance from models such as Zelinsky's (2008) "Target Acquisition Model" or Pomplun et al.'s (2003) "Area Activation Model." In these types of model, guidance is exclusively top-down: target- (template- or feature-) based. In fact, Zelinsky (2008) finds it "arguable whether a model that combines both top-down [target-template-based] and bottom-up [saliency] signals would be more successful than TAM in describing human behavior, at least in tasks in which the top-down target information [is] highly reliable" (p. 825). Such models, however, fail to address what determines target detection in search for (feature or feature conjunction) singleton targets, where there is no (reliable) target template to top-down guide the search (Müller et al. 1995; Weidner & Müller 2013); for example, is target "pop-out" based on a parallel attentive process operating over the whole display or a pre-attentive, salience-based process? One interesting possibility is that, on TP trials, detection decisions are triggered directly by the salience map – consistent with studies showing pop-out detection with no or minimal target identity processing (e.g., Müller et al. 2004; Töllner et al. 2012b) and some process of parallel distractor rejection taking place on TA trials (e.g., Müller et al. 2007). On more difficult search trials, the pre-attentive guidance mechanism could direct the attentive process to sample an area that surrounds the location of the highest salience. Here, models such as H&O's may indeed add to the traditional item-based models.

Chances and challenges for an active visual search perspective

doi:10.1017/S0140525X16000200, e150

Sven Ohi and Martin Rolfs

Bernstein Center for Computational Neuroscience & Department of Psychology, Humboldt University of Berlin, Berlin 10099, Germany.

sven.ohi@bccn-berlin.de

http://www.rolfs-lab.de/

martin.rolfs@hu-berlin.de

http://www.rolfs-lab.de/

Abstract: Using fixations as the fundamental unit of visual search is an appealing gear change in a paradigm that has long dominated attention research. To truly inform theories of search, however, additional challenges must be faced, including (1) an empirically motivated

definition of *fixation* in the presence of fixational saccades and (2) the biases and limitations of transsaccadic perception and memory.

In their target article, Hulleman & Olivers (H&O) argue for a conceptual change in characterizing visual search efficiency. The classical view explains visual search times as a function of the number of stimuli in a display (i.e., set size). According to the critique by H&O, this perspective constrains the study of visual search to a scenario that requires clearly defined objects viewed during prolonged fixation. Moreover, they argue, the traditional approach falls short of incorporating results from a larger range of search conditions – including overt visual search and searches in natural scenes in which items are not clearly defined. To overcome these limitations the authors present a theoretical framework that accounts for the number of fixations in a scene based on the assumption of an adjustable functional visual field (FVF), across which parallel processing takes place. In considering eye movements as a fundamental part of search, however, a number of challenges arise that, once faced, promise important theoretical insights for studies interpreted in this new context and beyond. We will focus on two challenges here.

1. What's a fixation? The authors' central aim is to understand search times based on the number of fixations during the search process. However, a fixation is not as clear-cut and discrete an entity as it might seem. Large primary saccades are frequently followed by smaller secondary saccades that often correct for errors in saccade landing position, but can also be observed after precise primary saccades (Ohl et al. 2011). Both primary and secondary saccades meet the criteria for a saccadic eye movement, but it remains unclear whether the interval between primary and secondary saccades should be considered an independent fixation. Moreover, even during instructed fixation, small microsaccades are observed at a rate of 1–2 per second (Rolfs 2009). Microsaccades have traditionally been considered *fixational* eye movements, suggesting that the interval between two microsaccades does not constitute an independent fixation. However, evidence accumulates that they are controlled by the same machinery as large saccades (Hafed et al. 2009; Rolfs et al. 2008) and fulfill the same purpose (Hafed 2011; Ko et al. 2010), namely, bringing a stimulus onto the part of the fovea that affords the highest resolution. Fixations separated by microsaccades, therefore, may need to be included when computing visual search times. This acknowledgment has two interesting consequences. First, the proposed framework might help clarify whether the intervals bordering on a microsaccade should be considered separate fixations. By comparing empirically observed numbers of fixations contingent on their definition (as either including microsaccades or not), future research could evaluate what definition of a fixation more accurately predicts the observed search times. Second, the presence of microsaccades during fixations may help resolve the dilemma that H&O face when explaining how search can be successful even in the absence of (large) saccades. Observers are not aware of their own microsaccades, and the generation of microsaccades has been linked to shifts of covert attention (Engbert & Kliegl 2003; Hafed & Clark 2002; Yuval-Greenberg et al. 2014). The perpetual execution of microsaccades results in more than one fixation even when observers are explicitly instructed to fixate while performing the visual search task. This variable number of fixations could be informative for characterizing covert visual searches and provide an opportunity to conceptualize it in H&O's framework.

2. Constraints of transsaccadic vision. Active vision is characterized by severe processing limitations that present challenges and constraints for theories of visual search. With each saccade, the incoming light reflected by an object will fall onto a new part of the retina and is thus processed by largely different neural populations in every retinotopic area in the visual processing stream. As a consequence, the visual system needs to keep track of the locations of relevant items as well as of their identities (see Cavanagh et al. 2010 for a review), including potential targets

and clear non-targets. There is strong psychophysical evidence that attended locations are updated across saccades (e.g., Jonikaitis et al. 2013; Rolfs et al. 2011), most likely relying on perisaccadic updating of visual priorities in visual attention-related brain areas (see Wurtz 2008 for a review). Indeed physiological results suggest that this updating of visual priorities (and, hence, the distribution of attention) involves the entire visual field, including distractor locations (Mirpour & Bisley 2012). Whereas this evidence suggests that the system is keeping track of the locations of potential targets and distractors, the accumulation of spatially disperse stimulus feature information across fixations has severe capacity limits. Indeed only three to four items are remembered correctly across saccades (Irwin 1991), and visual memory is heavily biased towards the saccade target (Bays & Husain 2008; Rolfs 2015). Thus far, H&O's framework considers a restriction only in how many visited locations can be remembered but does not take into account the visual system's limited ability to keep track of stimulus information across saccades. To the extent that this stimulus information is relevant for the search task, the bottlenecks that saccades impose on visual perception and memory fundamentally constrain the relationship between the number of fixations and an observer's search efficiency. Although item-based models do not address saccade-related constraints at all, the framework put forth by H&O provides a fertile ground to incorporate these insights from the study of active vision into the domain of visual search.

To conclude, research on human eye movement has revealed innumerable determinants shaping the alternating sequence of saccades and fixations—including their fundamental link to visual perception and memory. The framework presented by H&O provides a basis for the inclusion of these insights in the formulation of fixation-based theories of visual search. We highlighted two aspects—the controversial definition of fixations and the constraints imposed by transsaccadic vision—that provide challenges and opportunities for theories of active visual search.

Scanning movements during haptic search: similarity with fixations during visual search

doi:10.1017/S0140525X16000212, e151

Achille Pasqualotto

Faculty of Arts and Social Sciences, Sabanci University, Tuzla 34956, Istanbul, Turkey.

achille@sabanciuniv.edu

Abstract: Finding relevant objects through vision, or visual search, is a crucial function that has received considerable attention in the literature. After decades of research, data suggest that visual fixations are more crucial to understanding how visual search works than are the attributes of stimuli. This idea receives further support from the field of haptic search.

Finding objects in a complex environment is a crucial skill for the survival of humans and animals. Finding the suitable fruit in a cluttered forest, detecting predators disguised in the savanna, or simply finding a stapler on a cluttered office desk, all show the importance of the ability to find objects in an environment containing much irrelevant information (distractors). In most cases, and for humans in particular, this function is carried out by the visual modality, namely through *visual search*. Not surprisingly, visual search has been broadly investigated in humans and animals (Hickey et al. 2010; Proulx et al. 2014; Tomonaga 2007; Young & Hulleman 2013). Over the past several decades, theories of visual search have been constructed based on studies focusing on the number and the type of items used as distractors in visual search tasks (Duncan & Humphreys 1989; He & Nakayama 1992; Treisman 1982). In other words, the characteristics of the *items* were considered the central feature in visual search.

However, towards the end of the 90s, empirical data began to suggest that, rather than the characteristics of the items, another feature might better explain how visual search works. This feature is the number of fixations produced by participants during visual search tasks (Wolfe 1998b; Wolfe et al. 2010b; Zelinsky 1996). The more recent and mounting evidence in favour of the theory based on visual fixations prompted Hulleman & Olivers (H&O) to propose a novel theory to interpret past results and design future experiments. The authors conducted a meta-analysis to demonstrate how the number of fixations accurately predicts search times, and argued that the theory based on eye-fixations takes into account the physiology and the limits of the visual system. This led the authors to suggest that number of fixations would be better able to explain how visual search works than the number and type of the items. In fact, fixations are a constant factor (“fixations are fixations”), whereas items can be objects, faces, letters, numbers, and so forth, making experimental results highly task-dependent. This implies that studying fixation patterns would facilitate the comparison of results across different studies, and perhaps lead to a more comprehensive understanding of the underlying mechanisms of visual search.

This innovative approach is, however, limited to *visual* search. In real world, object finding could depend upon, and can be performed through, other modalities such as the *haptic* modality; for example, finding the keys in a pocket, finding a torch during a blackout, and so forth. The haptic modality consists of touch as well as proprioceptive and kinaesthetic cues (Gibson 1962; Heller 1984), and is very well suited to conveying information regarding shapes and positions (plus “extra” information such as temperature) of external objects (Lacourse & Fraszky 1997; Lederman & Klatzky 1987; Sann & Streri 2007). Interestingly, haptics and vision are broadly interconnected at the behavioural level (Ballesteros et al. 1998; Grabowecky et al. 2011; Newell et al. 2005; Pasqualotto et al. 2005; 2013a), and the two sensory modalities are strongly interlinked in the brain (Lacey et al. 2009; Pasqualotto et al. 2013b; Pietrini et al. 2004). Therefore, it seems reasonable to assume that, if visual search is based on visual fixations, haptic search might be based on hand/finger scanning movements.

Haptic search has been far less studied than visual search; nevertheless, both early and recent evidence suggests that scanning movements are as pivotal for haptics (Lederman & Klatzky 1987; Overvliet et al. 2007; Plaisier et al. 2008; 2009; Van Polanen et al. 2011) as fixations are for vision, thus supporting the assumption that the two modalities are strongly interlinked (see the previous paragraph). For example, in a haptic search task where participants had to find a target object among several distractors through active touch, haptic scanning patterns resembled patterns of eye-fixations observed in visual search (Overvliet et al. 2007). A similar study reported that, rather than the number and type of distractors, the strategy of haptic exploration (i.e., using one finger, one hand, or two hands) was the crucial factor that influenced haptic search performance (Overvliet et al. 2008). The same study also reported that systematic “zig-zag” haptic scanning patterns were observed across different searching conditions (e.g., different target objects), which suggests that scanning movements reflect a fundamental feature of haptic search. Furthermore, another study found that the “pop-out” effect well documented in the literature of vision (e.g., detecting a red object amongst blue objects) also occurs for haptics, and that haptic scanning movements were the best predictor for haptic search performance (Plaisier et al. 2008). The similarity between visual and haptic search is consistent with theories such as neural reuse (Anderson 2010) and metamodal brain (Pascual-Leone & Hamilton 2001), both of which suggest a substantial overlap in the brain across the areas processing input from *different* sensory modalities (Pasqualotto et al. 2016; Uesaki & Ashida 2015).

Although the novel approach to understanding the underlying mechanisms of visual search proposed in this article is supported by evidence from haptic search, the number of studies on haptic

search is still very limited compared to that on visual search, thus more research is needed to confirm those initial findings. In particular, it is critical to compare the patterns of visual fixations and haptic scanning movements arising within the same experimental setup *directly* to achieve a more holistic understanding of visual search as well as object search through other sensory modalities.

Mathematical fixation: Search viewed through a cognitive lens

doi:10.1017/S0140525X16000224, e152

Steven Phillips and Yuji Takeda

National Institute of Advanced Industrial Science and Technology (AIST), AIST Tsukuba Central 6-11, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan.

steve@ni.aist.go.jp yuji-takeda@aist.go.jp

https://staff.aist.go.jp/steven.phillips

https://staff.aist.go.jp/yuji-takeda

Abstract: We provide a mathematical category theory account of the size and location of the authors' Functional View Field (FVF). Category theory explains systematic cognitive ability via universal construction, that is, a necessary and sufficient condition for composition of cognitive processes. Similarly, FVF size and location is derived from a (universal) construction called a fibre (pullback) bundle.

Hulleman & Olivers (H&O) account for an impressively diverse array of visual search data with a single free parameter: the “size” of (number of items in) their putative Functional View Field (FVF). Nonetheless, we see two critical shortcomings: (1) FVF is purely descriptive and lacking independent motivation, which is indicative of an ad hoc assumption (Aizawa 2003); and (2) FVF size is potentially infinite in continuous domains, making it unclear how such cases are supposed to be unified with search in finite settings. In support of the target article, we provide a mathematical (categorical/topological) basis for FVF (size and location), called a fibre (pullback) bundle (Husemoller 1994), to help resolve these problems.

A category consists of a collection of objects, a collection of morphisms between objects, and an operation for composing morphisms (Mac Lane 1998). For example, every topological space is a category whose objects are the subsets constituting the topology, and morphisms are inclusions. Categories can model cognition by interpreting objects as cognitive states or spaces, morphisms as cognitive processes between states/spaces, and the operation as composition of cognitive processes. A universal construction is an arrangement whereby every morphism is composed of a common morphism and a unique morphism. This arrangement generalizes and refines classical/symbolic (Fodor & Pylyshyn 1988) and connectionist/functional (van Gelder 1990) compositionality by providing a necessary and sufficient condition for the indivisibility of certain clusters of cognitive capacities (Phillips & Wilson 2010; 2011; 2012), without the ad hoc assumptions that were problematic in other approaches (Aizawa 2003). Moreover, every universal construction is optimal in a certain category-theoretical sense. Thus, the preference for aligning relations over features in analogy (Gentner 1983), affording optimal transfer of source knowledge to target domain, derives from a universal construction (Phillips 2014). Visual search also involves compositionality and systematicity, hence our pullback approach to some differences between feature versus conjunctive search (Phillips et al. 2012). Similar considerations motivate our fibre bundle approach to FVFs.

We regard the FVF as a projection of visual information formalized as a fibre bundle (E, B, π, F) : a topological space E , called the total space, that is locally a product of base B and fibre F , together with a projection $\pi : E \rightarrow B$ that is a continuous surjective map. Projections can be filters, discussed in the target article, in the sense of maps from unfiltered display (total) spaces to filtered view (base) spaces. For example, the projection $\pi_1 : C \times O \rightarrow C$ filters out features in orientation space O so that attention is focused on features in colour space C —total space $C \times O$ is the product of base C and fibre O . Likewise, $\pi_2 : C \times O \rightarrow O$ filters out colour features to focus attention on orientation features, as the task demands.

Search involves changes in fixation that are bundle maps. In particular, a pullback bundle is obtained by “pulling back” a

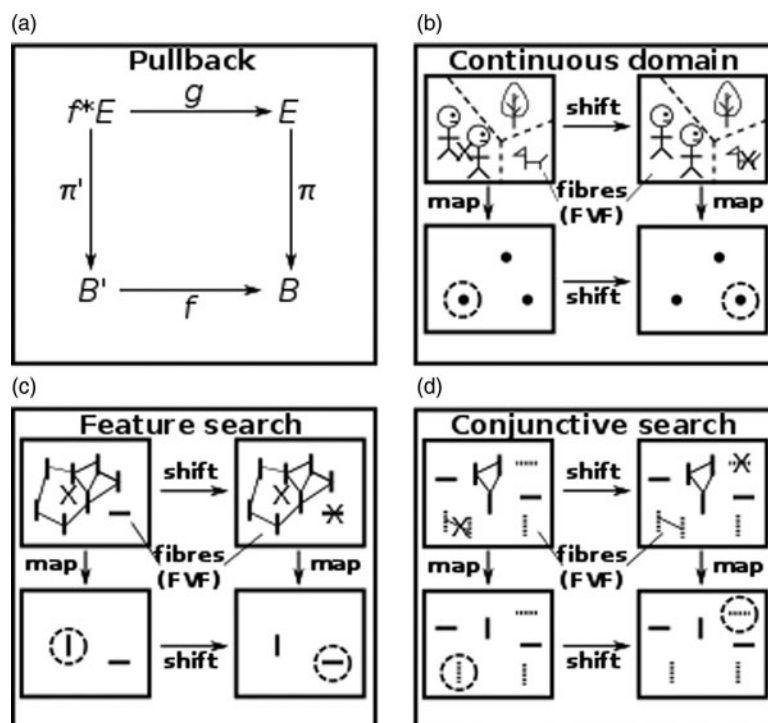


Figure 1 (Phillips & Takeda). Fibre bundle (a) commuting/pullback square, and corresponding examples for (b) a natural scene, (c) feature search, and (d) conjunctive search.

fibre bundle along a continuous map $f : B' \rightarrow B$ between base spaces, obtaining total space f^*B of pairs (b', e) , in a way that preserves bundle structure. That is, Figure 1(a) is a commuting (pull-back) square: $\pi(g(b', e)) = f(\pi(b', e))$. Commute means that fixation change after filtered view is the same as filtered view after fixation change, so search over view space is effectively search over display space. This construction is likened to a database lens, developed for a conceptually similar view update problem (Johnson et al. 2012), hence the expression “cognitive lens.”

FVF size and location are determined by the nature of the projection and an inverse. A fibre over a point $b \in B$, that is, the set of points in E that project to b , denoted $\pi^{-1}[b]$, corresponds to an FVF. Hence, the size of an FVF is the number of elements in $\pi^{-1}[b]$. A section of a fibre bundle is a continuous right inverse of its projection, that is, a function $\sigma : B \rightarrow E$ such that $\pi(\sigma(b)) = b$. The location of an FVF associated with point b in view space is the point $\sigma(b)$ in display space. The pullback condition (Mac Lane 1998) restricts bundles to disjoint unions of fibres, an optimal partitioning that prohibits gaps between fibres or overlaps.

Projections based on convex hulls of (topologically) neighbouring elements are one way to realize FVF for search in both natural and laboratory settings. For example, Figure 1(b) depicts a “natural” scene (E). Each convex hull (smallest set) enclosing one of the three scene components, that is, two people, a dog, and a tree, has a centre of mass, h_i . The projection (π) sends each point $e \in E$ to the closest centre. The fibre F_i is the region containing h_i , with boundaries indicated by dashed lines (cf. Voroni diagram). The base (B) is the (discrete) topological space on the three-centre set H , that is, the set of all subsets of H (cf. Delauney diagram). FVF location is the corresponding centre. An “X” indicates fixation before and after attentional shift, and dashed circles indicate corresponding items in the base. Pullback squares compose (Mac Lane 1998). So response time corresponds to the number of composed squares to termination. By commutativity, search need only involve the three items in the base, rather than a very large (potentially infinite) number of locations in the display; greater resolution implies more locations (cf. texture-based search).

An analogous situation applies to feature versus conjunctive search, shown in Figure 1(c, d). Lines connecting bars indicate neighbours in topological space; connected graphs correspond to fibres, which are larger in feature than conjunctive search, hence feature search is generally more efficient. Off-item fixation corresponds to a virtual bar at the “centre of mass” of a multinode graph, so fixation need not coincide with a displayed bar. This situation is akin to perceptual grouping (Duncan & Humphreys 1989). Similar considerations apply to conjunctive search (Wolfe et al. 1989): Nearby items are less likely to be of the same kind, hence FVFs are smaller and the number of fibres greater, implying the observed steeper search slope. A base can contain non-visual items, for example, categories affording category-guided search (Zelinsky et al. 2013). The challenge is to develop an FVF model incorporating the mathematical theory.

Task implementation and top-down control in continuous search

doi:10.1017/S0140525X16000236, e153

Wolfgang Prinz

Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig 04103, Germany.

prinz@cbs.mpg.de

<https://www.cbs.mpg.de/staff/prinz-10359>

Abstract: Evidence from continuous search suggests that targets are detected by default, whereas distractors are processed in considerable depth. These observations shed light on task implementation and top-down control. Task implementation builds on forming dynamic

distractor models, based on continuous integration of distractor-related information. Top-down control builds on using these models for testing upcoming stimulus information.

The target article claims that the analysis of visual search performance must proceed to larger functional units, moving from items to fixations. This commentary seconds and extends this claim, drawing on earlier studies of active, continuous search. By shedding new light on mechanisms of target detection and distractor processing, these studies provide complementary evidence about task implementation and top-down control in search.

In continuous search tasks participants are exposed to large arrays of items whose scanning requires extended fixation sequences. Item arrays may be arranged in slim columns through which the eye wanders from top to bottom (e.g., Neisser 1963) or rectangular blocks that are scanned line-by-line, as in reading (Prinz 1986). Items may be letters, digits, or other arbitrary elements. Typically, the task requires scanning through the array for a pre-specified target, and the scan terminates upon its detection. A typical scan will thus encounter a large number of distractors before it eventually comes across the searched-for target.

Performance in such tasks seems to call for a simple and intuitive control model, claiming that searching for something requires forming and maintaining a selective attentional set that highlights target-related information. The searcher may, for example, be seen to form a template of the target and maintain it in a primed, pre-activated condition. With this template in mind, she will then scan the search array until she encounters an item that matches it. According to this view, search is controlled by target-related information and target detection is accomplished through target identification. This seemingly natural view was in fact entailed in Neisser’s model of search control (Neisser 1967). Based on Selfridge’s Pandemonium architecture (Selfridge 1959), the model claimed that task implementation activates stimulus analyzers for targets, but not for distractors, yielding deep target processing but only shallow distractor processing. As a result, distractors remain unidentified whereas target detection is brought about by target identification.

However, there is now substantial evidence suggesting that this view is fundamentally mistaken. The evidence applies to both target detection and distractor examination (for overviews, see Prinz 1977; 1986).

First, target detection does not require deep modes of target processing. Targets are rather detected without being identified. This conclusion is suggested by several converging observations. (1) Pure detection: In tasks requiring parallel search for several targets simultaneously, participants may detect upcoming target locations without/before knowing which target is actually present. Targets can thus be detected and localized before they can be identified. (Neisser 1963; 1967, p. 100). (2) Pseudotarget detection: When a novel distractor appears in the search array, it is often treated as a target. Novel distractors are thus detected without being searched for (Prinz 1979, Exp. 1; Prinz et al. 1974). (3) Detection at distance: Targets are often detected prior to being fixated. They can thus be detected without being the focus of spatial attention (Prinz 1983; Prinz & Kehler 1982). In sum, these observations suggest that targets are detected by default, not by identification. Target detection seems thus to rely on representational resources other than selectively primed target templates.

Second, distractor examination does not rely on shallow modes of distractor processing. Distractors are rather processed at considerable depth. This conclusion is suggested by a further set of observations: (1) Distractor complexity: Array complexity (as defined by the size of the distractor set) is one of the most powerful drivers of search performance (Prinz 1979). (2) Distractor redundancy: Sequential redundancy of distractor strings and arrays has been shown to drive performance as well. Importantly, distractor redundancy is automatically detected and exploited (Prinz 1979). (3) Distractor transfer: In paradigms requiring

switches between search tasks, substantial transfer is obtained when the tasks share the same distractors (Prinz & Ataian 1973). Remarkably, this also applies to visually different distractors that share the same names (Prinz et al. 1972). In sum, these observations suggest that distractor-related information is not discarded at all. Instead, it seems to be automatically identified within fixations, updated and integrated across fixations, and stored across consecutive trials.

Taken together, these findings suggest a framework for task implementation and top-down control that is actually the inverse of Neisser's intuitive model (Prinz 1977; 1986; for overviews of more recent evidence on information integration and task-driven control see Schneider, 2013; Schneider et al. 2013). This framework builds on three major claims: (1) Incoming information is continuously processed and integrated over space and time (i.e., within and across fixations, respectively). (2) Integration processes generate, maintain and update a dynamic forward model of current and impending distractor information (simultaneously for features, items, strings, etc.). (3) Once such a model is in place subsequent fixation samples are tested against it. As long as they match it, the scan is continued (= pure distractor samples). However, when a sample fails to match it, the scan gets disrupted for closer examination (= mixed distractor/non-distractor samples).

Regarding target processing, this framework accounts for detection by default. Targets are oddballs that fail to match the dynamic distractor model, be it at feature, item, or string level. Regarding distractor processing, it accounts for task implementation and top-down control. Task implementation arises as a by-product of automatic distractor processing: Continuous integration of distractor information within and across fixations generates a dynamic distractor model that instantiates the control structure for the current task. Top-down control is then based on using this model for testing upcoming fixation samples.

These are lessons from continuous search. Why should we take them to heart? On the one hand, continuous search differs from discrete search in several important respects so that only some of these lessons may be directly applicable to that domain. On the other hand, continuous search must be considered a powerful paradigm in its own right. It addresses processing mechanisms for active, extended search as they underlie natural search episodes in real-world scenes and settings.

Those pernicious items

doi:10.1017/S0140525X16000248, e154

Ruth Rosenholtz

Department of Brain & Cognitive Sciences, CSAIL, Massachusetts Institute of Technology, Cambridge, MA 02139.

rruth@mit.edu

<http://persci.mit.edu/people/rosenholtz>

Abstract: Hulleman & Olivers (H&O) identify a number of problems with item-based thinking and its impact on our understanding of visual search. I detail ways in which item-thought is worse than the authors suggest. I concur with the broad strokes of the theory they set out, and also clarify the relationship between their view and our recent theory of visual search.

Our impression of a scene usually includes objects and their properties. When crossing the street, we consider the location and speed of a nearby car. However, just because we recognize “things” at the output of perception and employ high-level reasoning about objects, this does not mean that our visual systems operate on presegmented things. This is a common and tempting cognitive error, which can hamper uncovering the true mechanisms of vision.

Objects make for a useful abstraction. It is natural, therefore, that many theories of vision describe processes as operating on

objects and their features. For example, preattentive vision has been depicted as encoding *item* locations and features. (This is distinct from knowing *image* features, such as the outputs of V1 cells.) According to this view, search is slow because serial selective attention is necessary to bind those features together. Such word models enable easy intuitions and guide new experiments. Furthermore, abstracting from the image input to things and their features can sometimes make modeling tractable, as with signal detection theory.

The authors argue that a focus on items has tainted our ideas about search: it has hampered understanding search in real-world images, for which the set size (number of items) is ill-defined (Rosenholtz et al. 2007; Wolfe et al. 2011a; Zelinsky 2008); it has led to a focus on selective attention as the limiting mechanism, discounting the role of eye movements; it caused the field to focus on the easy end of the performance spectrum; and it has led to over-estimation of the importance of item location. I would argue that thinking about items is even more pernicious.

Item-based theories have not merely biased our choice of stimuli by limiting use of real-world images. Experimenters often design stimuli to preserve the preeminence of the item. One must avoid alignment, which might produce perceptual groups, or else risk violating assumptions that the items can be treated independently. This is analogous to visual short-term memory experiments that seem designed to give the subject little choice but to remember *items*; it should not surprise us when slot models do well.

Relatedly, only a handful of experiments have studied the effects of image transformations on search: What if we make the items larger or the bars thinner, change the sign of contrast, add noise to the display, or make the displays more dense? Unless these transformations interfere with item visibility, none should have an impact if items are the atoms of search. Yet there is evidence that such transformations do have a significant effect (e.g., Beck et al. 1987; Chang & Rosenholtz 2014; Graham et al. 1992; Rubenstein and Sagi 1996).

More broadly, it is risky to think of the visual input as consisting of an array of items with particular experimenter-defined features. Vertical rectangular bars also contain horizontal edges; oblique filters will also respond to those bars; the “white space” between items also has features; and some features of the display may have a scale larger than any individual items.

The dominance of item-based theories has led to a serious disconnect between theories that essentially operate on experimenter-labeled stimuli (items and their nominal features) and those that operate on actual images. In working with real images, a number of reasonable search strategies do not require items as such, for example applying a template throughout the image and looking for locations with a strong response (Zelinsky 2008). If one does attempt to implement item-based theories, it quickly becomes clear that neither segmenting the items nor determining their supposed “features” is trivial. One is left with the puzzle of why one is “allowed” to use bound features to “pre-attentively” segment the image into items, but not to recognize the target.

It is hard not to think in terms of items. Despite their main thesis, Hulleman & Olivers (H&O) suggest that target-distractor discriminability is important for setting the size of the functional visual field (FVF). Why discriminability of the *items*? This leaves the puzzle of why search asymmetries abound, as surely target-distractor discriminability is generally the same as distractor-target discriminability. We argue that the major determinant of search performance is crowding (*not* retinal resolution), which demonstrates that peripheral vision operates over sizeable patches, typically containing multiple items. Discriminability of patches is what matters in this scheme (Rosenholtz et al. 2012b).

The authors allude to one theory attributing crowding to limited attentional resolution (Intriligator & Cavanagh 2001). This is subtly item-centric, presuming that attention aims to select only

a single item. We have argued this is not ideal in real images (Rosenholtz & Wijnjes 2014). Other theories of crowding also describe mechanisms that operate on items (Greenwood et al. 2009; 2012; Levi & Carney 2009; Parkes et al. 2001; Pöder and Wagemans 2007; Strasburger 2005; van den Berg et al. 2012). Recently we have shown that a number of results used to test these item-based theories can instead be explained by the information available in a rich set of image statistics (Keshvari & Rosenholtz 2016). These same statistics plausibly underlie scene perception (Ehinger & Rosenholtz 2016; Rosenholtz et al. 2012a), suggesting a single encoding scheme could both extract the scene context and support search, in agreement with H&O.

The target article presents clever and thoughtful critiques of prevailing theories, and a new model. The parallels to recent work in my lab are fairly clear (Rosenholtz et al. 2012b; Zhang et al. 2015), though differences raise important questions. We agree that search likely involves parallel processing, punctuated by serial shifts of the point of fixation. Peripheral vision limits the information available at a glance. Our view is that, rather than being a mechanism, the FVF might *describe* the more informative image regions. It would degrade smoothly, with some regions providing more information than others. It need not be continuous; eccentric uncrowded regions might provide more information than closer crowded ones. The authors are somewhat unclear on these points: Does the FVF have hard edges, outside of which no information exists for telling apart the target and distractor? Is it mechanistic or descriptive? Does some mechanism set the size? If so, how, and why?

What fixations reveal about oculomotor scanning behavior in visual search

doi:10.1017/S0140525X1600025X, e155

Zhuanghua Shi,^a Xuelian Zang,^{a,b} and Thomas Geyer^a

^aGeneral and Experimental Psychology, Department of Psychology, LMU Munich, 80802 Munich, Germany; ^bChina Centre for Special Economic Zone Research, Shenzhen University, Guangdong Sheng 518060, China.

strongway@psy.lmu.de zangxuelian@gmail.com
geyer@psy.lmu.de

Abstract: Hulleman & Olivers' (H&O's) conceptual framework does not consider variation of fixation duration and its interaction with the size of the functional viewing field (FVF). Here we provide empirical evidence of a dynamic interaction between the two parameters, suggesting that fixations, as the central unit in H&O's framework, should be studied on both the spatial and temporal dimensions.

By taking fixations, not individual items, as the central unit, Hulleman & Olivers (H&O) put forward a promising, unified account of both eye movements and manual reaction times (RTs) in visual search. However, their conceptual framework makes two oversimplified assumptions: (1) the size of the functional viewing field (FVF) being solely dependent on the visual discriminability of the search elements; and (2) constant FVF processing time (i.e., a constant fixation duration of 250 ms), ignoring any dynamic interactions between the two parameters. Although the assumption of constancy of fixation durations makes the framework easily comparable with traditional, item-based selection models, it limits the explanatory potential of H&O's account, as we will outline in this commentary.

It is generally accepted that “fixate” and “move” oculomotor activities are governed by parallel “when” and “where” commands generated across the entire visual-perceptual hierarchy (Findlay & Walker 1999). Concerning top-down influences, fixation durations are influenced by task difficulty (Hooge & Erkelens 1998; Moffitt 1980; Pomplun et al. 2013), memory about spatial context (van Asselen et al. 2011; Zang et al. 2015), visual search strategy

(Geyer et al. 2007), and multisensory experience (Zou et al. 2012). For example, Geyer et al. (2007) compared fixation durations between static and dynamic search displays with identical target-distractor discriminability, except that search items were randomly reshuffled every 117 ms in the latter condition. Mean fixation duration, as well as the latency of the first saccade, was increased by some 100–150 ms for the dynamic compared to the static condition, although “standard” measures of search efficiency (slope of the search function) were comparable between the two types of display. These findings clearly suggest that fixational dwell times are not solely under the control of the current sensory environment, or in H&O's terms, the perceptual discriminability of the search items. Instead, observers' strategic efforts in solving the task at hand must also be considered in accounting for such extended fixation durations (Geyer et al. 2007).

Rather than being independent, in most cases fixation duration and the FVF interact in a nonlinear fashion (Nuthmann et al. 2010; Unema et al. 2005). One strong piece of evidence of a dynamic interaction between the two parameters comes from an oculomotor study on the “pip-and-pop” effect (Zou et al. 2012). In “pip-and-pop” visual search displays, beeps are synchronized with (task-irrelevant) color changes of the target, which is presented in a cluttered and heterogeneous item field (with search being extremely “inefficient”). Zou et al. found that fixation durations increased by some 150 ms for beep-present versus beep-absent trials: an “oculomotor freezing” effect. Such extended fixations at beeps allow information to be sampled over a larger FVF, as indicated by larger saccade amplitudes immediately after the beeps. In other words, beep-induced prolonged fixation times and subsequent large saccade amplitudes mediate fast detection of target presence, yielding the “pip-and-pop” effect. This pattern also suggests that the oculomotor scanning strategy can affect the rate of information processing, as evidenced by increased information uptake per fixation for the beep-present relative to the beep-absent condition. Another very recent study (Zang et al. 2015) on context-based guidance of visual search also revealed a beneficial effect of extended fixation duration on task performance. In this study observers were first trained with an artificial FVF size, implemented by a gaze-contingent tunnel-viewing technique. With 4–5 items visible inside of the FVF, the mean fixation duration was already extended in the training session for repeated “old,” compared to randomly generated “new,” display (item) layouts. Further, the scan path for old relative to new displays was closer to the optimal scan path, indicating that learned context improves the efficiency of oculomotor scanning. Increased fixational dwell times and shortened scan paths for old relative to new displays remained evident even after the constraining tunnel view was removed from the task. Such dynamic adjustments of fixation duration and saccade amplitude are quite common during scene search. It has been shown, for instance, that fixation duration and saccade amplitude gradually change over the first few seconds, and then approach their asymptotic levels (Unema et al. 2005). Both asymptotes, however, depend on the number of objects in the scene, which indicates that the complexity of the scene, too, changes oculomotor scanning.

These findings, amongst others, provide converging evidence that the size of the FVF and fixation duration are not determined by visual discriminability alone, as assumed by H&O. Rather, oculomotor scanning is dynamic in that the size of the FVF and fixation duration must be considered together to discern moment-by-moment adjustments of information processing. Despite the H&O conceptual framework's current lack of flexible oculomotor parameters, the idea of fixation as a central processing unit of visual search remains very promising. However, to incorporate the above findings of dynamic interactions between fixation duration and saccade amplitude, we propose that fixational eye movements are best characterized by both spatial (i.e., the size of FVF in H&O terms) and temporal (i.e., fixation duration) factors. Combining the two could provide insight into how oculomotor scanning strategies influence the fixation-by-fixation information

processing rate, which might turn out to be the distinguishing feature for comparing different visual search tasks.

Item-based selection is in good shape in visual compound search: A view from electrophysiology

doi:10.1017/S0140525X16000261, e156

Thomas Töllner^{a,b} and Dragan Rangelov^a

^aDepartment of Experimental Psychology, Ludwig-Maximilians-Universität München, 80802 Munich, Germany; ^bGraduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany.

thomas.toellner@psy.lmu.de

<http://www.psy.lmu.de/exp/people/ma/toellner/index.html>

rangelov@psy.lmu.de

<http://www.psy.lmu.de/exp/people/ma/rangelov/index.html>

Abstract: We argue that although the framework put forward by Hulleman & Olivers (H&O) can successfully explain much of visual search behaviour, it appears limited to tasks without precise target identification demands. In particular, we contend that the unit of selection may be larger than a single item in standard detection tasks, whereas the unit may mandatorily be item-based in compound tasks.

The target article offers an exciting new perspective on how the cognitive system samples visual input for relevant information. Although we are in agreement with this account with regard to detection tasks, our commentary focuses on a potentially crucial limitation: *generality across task settings*. Despite Hulleman & Olivers' (H&O's) endeavour to highlight similarities between different search tasks, we argue that this issue deserves more attention. In particular, we contend that item-based selection is likely to be mandatory in tasks that require precise, focal-attentional stimulus analysis. Here we discuss three experimental findings that challenge the notion that the search process is equivalent in detection and compound tasks.

As noted by H&O, some of our findings indeed suggest that attentional selection is similar across tasks (Exp. 1; Töllner et al. 2012b). This view is supported by the activation profile of the Posterior Contralateral Negativity (PCN, or N2pc) – an ERP wave generally agreed to reflect attentional selection (Eimer 1996; Luck & Hillyard 1994; Töllner et al. 2012a) – which was elicited equally for search tasks in which observers had to detect, localize, or identify physically identical targets. However, this pattern was restricted to conditions with relatively high target prevalence (between 66.6% and 100%). In the same study (Exp. 2), we observed a strong influence of target prevalence on attentional selection: Using a detection task, PCN amplitudes increased gradually with decreasing target frequency (80% > 50% > 20%). Following one proposal according to which the PCN may represent the target's saliency/priority signal on the attention-guiding “master map” (Töllner et al. 2011; 2015a), this amplitude modulation may result from a cortical amplification mechanism (Egner & Hirsch 2005) that comes into play when searching for low-prevalence targets. In particular, the visual system may adapt to reduced target probabilities by boosting incoming relevant information through a context-sensitive amplifier that keeps track of environmental statistics. This biasing of target signals may automatically translate into enhanced activations at the next processing stage – the attention-guiding master map (likely represented by the PCN) – thereby raising detectability of rare targets to the level of frequent targets. Another factor that may have contributed to this amplitude gain is refractoriness (Woods et al. 1980). It is conceivable that neural populations associated with the coding of attentionally selected targets are more prone to refractoriness in conditions of 80% target prevalence in which the average inter-target interval was 2.5 s, as compared with

20% prevalence with, on average, a 10-second interval. Whatever the exact interplay of mechanisms, findings of attentional selection being frequency-dependent have implications for H&O's framework. Because a target is present on every trial in the typical compound task but only on half of the trials in the standard detection task, we surmise that attentional selection is not directly comparable among different search tasks using variable target frequencies.

Furthermore, H&O propose that compound-search tasks involve a cascade of processing steps, with sequential reductions of the functional visual field (FVF) to meet the requirements of focal-attentional analysis: the higher the requirements, the smaller the FVF. Critically, H&O assume that the first stage of these processes is similar across tasks, and that the initial size of the FVF depends exclusively on target discriminability. Although we agree with the general idea, we doubt that the initial stage of target selection is entirely detached from the subsequent target identification stage. Our scepticism derives from recent EEG evidence (Mazza et al. 2007; Rangelov et al. 2013b; Töllner et al. 2013) demonstrating that target identification involves a post-selective process that extracts detailed object information from working memory (WM). Specifically, we observed the CDA (or SPCN) wave – a well-established ERP marker of the maintenance of (Vogel & Machizawa 2004; Wiegand et al. 2014) but also the access to (Töllner et al. 2014; 2015b) WM representations – being elicited exclusively in target identification tasks, but not in tasks requiring simple target detection or localization. That target identification is carried out in WM may consequently place further constraints on the number of items participants may select in parallel in the first place. Given that WM is limited in capacity to about three to four items (Cowan 2001), it would appear plausible that participants strategically adjust the size of their FVF in relation to their available WM capacity. Note that this consideration is strikingly supported by the fact that three to four objects match exactly the target number at which both the PCN (Mazza et al. 2013) and the CDA amplitudes (Vogel & Machizawa 2004) reach asymptotic level. It is as if the stage of attentional selection (PCN) were naturally tuned to take in only an amount of information that can be fairly handled at subsequent WM stages (CDA). Moreover, no such set size modulations of the PCN are evident if the task requires simple detection, as opposed to exact enumeration (in WM), of multiple targets (Mazza & Carrazza 2011). In the light of these findings it appears unlikely that the initial stage is identical across tasks. Instead, we argue that, while in standard detection tasks the unit of selection may well be set according to target discriminability, in compound tasks the unit is additionally constrained by individuals' capacity to evaluate target information in WM. The unit of selection may even scale down to a single item; for example, when the task requires precise focal-attentional analysis in the face of highly similar distractors.

Finally, investigations of search slopes indicate qualitative differences between detection and compound tasks. Specifically, in detection tasks reaction times to pop-out targets remain constant across different display densities (Young & Hulleman 2013), whereas reaction times to the same targets decrease as more distractors are added in compound tasks (Bravo & Nakayama 1992; Rangelov et al. 2013a). In H&O's framework, search slopes are determined by the FVF size, which, in turn, depends on target discriminability. Consequently, FVF size should be the same for similar targets. As negative search slopes indicate, however, this is not the case.

The FVF might be influenced by object-based attention

doi:10.1017/S0140525X16000327, e157

Poutasi W. B. Urale

School of Psychology, The University of Auckland, Auckland 1010, New Zealand.

crimsonshinobi@gmail.com

Abstract: Hulleman & Olivers (H&O) argue that the primary unit in search should be fixations, and in doing so posit a Functional View Field (FVF). There is evidence from the object-based visual attention literature that the FVF may not process visual information uniformly. Here I sketch how object-based attention may influence processing *within* the FVF as well as the *shape* of the FVF.

Hulleman & Olivers (H&O) have given a deeply convincing case for why we should consider *fixations*, and not items, as the fundamental unit of search. Central to their framework is the Functional View Field (FVF), which is an area of the visual field centred on fixation, where an item can be expected to be detectable. Towards the end of their article, H&O pose the question of what factors determine the size of the FVF. One important question they did *not* ask is whether the FVF can change shape, and if so, under what circumstances. Moreover, H&O leave open the possibility of grouping mechanisms *within* the FVF. Here I consider how findings relating to object-based attention provide possible answers to these questions.

To begin, it is accepted that visual attention has both spatial- and object-based components (Soto & Blanco 2004). Spatially based visual attention involves attention being directed to a general *area* of the visual scene (Downing 1988). Object-based visual attention (OVA), on the other hand, is directed toward objects or groups of elements adjoined by Gestalt principles (Neisser 1967), and it has been supported in part by findings that, all else being equal, RTs are faster when a cue and target appear within the same object compared to when they are not (Egley et al. 1994). While there is an obvious spatial component to visual search, it is interesting to consider how, in the context of H&O's framework, the FVF may process information given what is currently known about object-based attention.

Evidence for intra-FVF grouping can be tentatively drawn from studies using the aforementioned cued-rectangle paradigm. For example, Norman et al. (2013), report an object-based effect with rectangles that were not conscious to the viewer. Viewers fixated on a point, then made a timed response to a target stimulus preceded by a cue. As with previous findings using this paradigm (e.g., Egley et al. 1994), viewers were faster to detect the target when it appeared within the same object as the cue. Applied to H&O's framework, this would suggest that once a fixation is made at a given location, such as on a cue, the FVF's contents are not processed uniformly (as H&O readily admit); rather, they are subject to grouping mechanisms that prioritise some visual elements within the FVF over others.

There is, however, disagreement over the mechanisms of object-based attention. Some argue that several OVA effects exemplify *attentional prioritization* of locations that simply assigns a higher probability of the target appearing within the boundaries of an attended object (Shomstein & Yantis 2002; Greenberg et al. 2015), not unlike priority of scene locations based on local contrast (e.g., Wolfe 2010). Thus, any possible grouping in the visual scene may facilitate search, but only because it guides fixations, not because of any alteration of the FVF. As a result my interpretation of Norman and colleagues' (2013) findings remains tentative, as they did not account for eye-movement during their task.

How might OVA influence the *shape* of the FVF? One way is by sensory enhancement (Desimone & Duncan 1995). Imagine that the "default" FVF during visual search of a given difficulty is a circular patch with fuzzy boundaries. When there are no sufficiently strong grouping factors between visual elements, such as in most search arrays, only the contents of the circular FVF are processed. However, if a peripheral visual element is grouped with an element within the "default" FVF, the signal from the peripheral visual element is "boosted" and is subsequently processed more thoroughly than it would have been otherwise, effectively expanding the FVF. Wannig et al. (2011) provide evidence to support such an account. They recorded two receptive fields (area V1) of monkeys during a visual task. The monkeys were trained to foveate on a target line stimulus, and during the task the target

line was flanked by task-irrelevant line stimuli. After fixation on the task-relevant stimulus, the authors found an increase in activity from the receptive field corresponding to a task-irrelevant line stimulus, but *only* if the task-irrelevant stimulus was *collinear* with the task-relevant stimulus. This demonstrates that: (1) attentional spreading takes place in early representations of visual elements grouped by Gestalt factors; (2) this attentional spreading occurs automatically (i.e., attention spreads to task-irrelevant stimuli). Further experimentation will be needed to see whether this kind of attentional spread has any bearing on visual search.

H&O remarked that their framework has no problem explaining search across a messy scene bereft of clearly-defined items. I agree, but I also stress that there is still much room within this framework for a scene to be divided into object-like groupings during search. It would be out of neglect for what is known about object-based attention to assume that the FVF always processes information irrespective of object groupings. Furthermore, although retinal physiology is no doubt crucial in determining the boundaries of the FVF, we have good reason to believe that these boundaries are also subject to early "signal boosting" as a result of Gestalt grouping.

Looking further! The importance of embedding visual search in action

doi:10.1017/S0140525X16000273, e158

John van der Kamp^a and Matt Dicks^b

^aVrije Universiteit Amsterdam, 1081 BT Amsterdam, The Netherlands;

^bUniversity of Portsmouth, Portsmouth, Hampshire PO1 2ER, United Kingdom.

j.vander.kamp@vu.nl

<https://vu-nl.academia.edu/Johnvanderkamp>

matt.dicks@port.ac.uk

[https://researchportal.port.ac.uk/portal/en/persons/matt-dicks\(c8460293-9bb8-43a1-bdb1-18f8ff8af9f9\).html](https://researchportal.port.ac.uk/portal/en/persons/matt-dicks(c8460293-9bb8-43a1-bdb1-18f8ff8af9f9).html)

Abstract: A unified account of visual search in complex everyday environments requires additional deliberations on the functional grounding of Hulleman & Olivers' (H&O's) functional viewing field (FVF) model. Their model can accommodate exploitation of information that is distributed across the immediate environment. Yet the differences in search between genuinely interacting in the environment and merely watching it should challenge researchers to look further.

The primary function of visual search is to gather information for guiding actions with persons, objects, and events in everyday environments. For example, a pedestrian walking down a busy city street, and a police officer chasing a suspect are examples of complex everyday environments that require people to continuously search information to act adaptively. Comparably, during expert search, a soccer goalkeeper trying to stop a penalty kick scans the penalty taker's body (e.g., the eyes, hips, legs, feet) for information that guides the decision of when and where to move.

In contrast to everyday environments, participants in a typical visual search research paradigm are seated and required to identify, as quickly as possible, an L among T's presented on a monitor. It is unlikely that this seated-monitor paradigm does justice to the intricacies of visual search in complex environments; this may be equally true for the associated models that presume that visual search proceeds on the basis of individual items (Treisman & Gelade 1980). Conversely, Hulleman & Olivers' (H&O's) FVF model takes the fixation (and thus eye movements) as the central unit of visual search. This is an important step because it opens up the possibility to go beyond research paradigms that only require participants to (passively) watch screens. In particular, H&O's perspective potentially permits researchers to explicate the observed pick-up of global information that occurs when searching complex everyday environments (Diaz et al. 2012). However, to fully capture the richness of visual search in

complex everyday environments, H&O's model would benefit from a thorough consideration of the *functional* characteristics of visual search.

If visual search functions to gather information for guiding our interactions with the environment, where the information resides and what action the information controls is of key concern. Gibson (1979) asserted that information resides in structured energy arrays that surround us. For example, the optic array is the surrounding light patterned by reflection against the surfaces, objects and persons (including the observer!) of the environment. Hence, the structured light patterns are specific to, and inform about, the environment. Accordingly, perception is the pick-up of this information in the optic array. Thus, it is crucial to recognize that an observer always moves, even if these movements would be restricted to the eyes! The information the observer exploits is a continuous flux, but within these unceasing transformations some patterns remain unchanged or invariant. Broadly speaking, Gibson proposed that invariances specify the (unchanging) environment, while the changes specify how the observer relates to the environment; they inform the observer about the actions that the environment affords. This implies that the observer's body, head, and eye movements co-structure information for guiding the observer's interaction with the environment, and thus must be part of any unified account of visual search.

To illustrate let us return to the soccer goalkeeper trying to stop a penalty kick. Typically, the ball moves at a speed that leaves a goalkeeper insufficient time to decide which side to dive on the basis of ball flight information. Therefore the goalkeeper must anticipate the direction of the dive based on information that resides in the penalty taker's movements. Expert goalkeepers distinguish themselves from their less successful counterparts in how they visually search the penalty taker's body for gathering this information. They make a small number of fixations of longer duration to fewer locations (Savelsbergh et al. 2002). Intriguingly, they particularly make long fixations on the empty space *in between* the non-kicking leg and the ball instead of making a sequence of fixations between different locations (Piras & Vickers 2011). This finding concurs with analyses that the most reliable information is distributed across different body locations rather than being located at one joint or body part (Diaz et al. 2012). H&O's FVF model can easily accommodate these observations. Within the model the FVF is defined as "the area of the visual field around fixation from which a signal can be expected to be detected" (sect. 5.1, para. 2). Importantly, the field is not fixed but varies in size. The smaller the field, the more fixations are needed and the more time the observer needs to search. Accordingly, expert goalkeepers may have a larger FVF than less skilled goalkeepers, allowing them to exploit the distributed information with less extensive visual search. This skilled search behaviour potentially provides experts with more reliable and timely information for ball interception.

H&O's perspective is largely limited to (typical) seated-monitor paradigms, which address how eye movements are used to search the environment. Yet in complex everyday environments, eye movements are but one means of gathering information, as a person's search also relies heavily on head and (whole) body movements. The over-reliance on seated-monitor experimental tasks leads to a limited view of visual search and may especially obscure its functional aspects. Perception and action, even in an *identical* environment, exploit different information and induce different patterns of visual search (Van Doorn et al. 2009). In this respect, we have previously reported that a soccer goalkeeper's visual search is fundamentally different when watching a penalty taker on a screen and verbally predicting kick direction compared to when actually facing a penalty taker in real-time and attempting to intercept the ball (Dicks et al. 2010). Perhaps surprisingly, on the pitch less time appears to be spent searching the body, while fixation towards the ball increases. It is likely that the different functional requirements affect the spatio-temporal structure of visual search in many more ways than a change in

the magnitude of the FVF. A crucial challenge for any account of visual search, including the FVF model, is to spell out in more detail how functional requirements systematically affect the eye, head, and body movements for gathering information in complex everyday environments. To this end, looking further than monitors is a necessity!

Don't admit defeat: A new dawn for the item in visual search

doi:10.1017/S0140525X16000285, e159

Stefan Van der Stigchel^a and Sebastiaan Mathôt^b

^aDepartment of Experimental Psychology, Helmholtz Institute, Utrecht University, 3584 CS Utrecht, The Netherlands; ^bAix-Marseille University, CNRS, LPC UMR 7290, Marseille, 13331 Cedex 1, France.

s.vanderstigchel@uu.nl

<http://www.attentionlab.nl>

s.mathot@cogsci.nl

<http://www.cogsci.nl/smathot>

Abstract: Even though we lack a precise definition of "item," it is clear that people do parse their visual environment into objects (the real-world equivalent of items). We will review evidence that items are essential in visual search, and argue that computer vision – especially deep learning – may offer a solution for the lack of a solid definition of "item."

To say that items do not play a role in visual search is to admit defeat. Even though we lack a precise definition of "item," it is clear that people do parse their visual environment into objects (the real-world equivalent of items in visual search). In this commentary, we will review evidence that items are essential in visual search; furthermore, we will argue that computer vision – especially deep learning – may offer a solution for the lack of a solid definition of "item."

In the model of Hulleman & Olivers (H&O), search proceeds on the basis of fixations that are used to scan a visual scene for a target. Although we appreciate its parsimony, the model lacks a crucial aspect of visual search: the decision where to look next. The model simply assumes that an arbitrary new location is selected. Yet there is abundant evidence that fixation selection is not random but rather results from integration of top-down and bottom-up influences in a common saccade map (Meeter et al. 2010; Trappenberg et al. 2001). That is, we look mostly at things that are salient or behaviorally relevant (Theeuwes et al. 1998) and, crucially, behavioral relevance is related to how we parse visual input into items (e.g., Einhäuser et al. 2008). Consider repetition priming: People preferentially look at distractor items that resemble previous target items (Becker et al. 2009; Meeter & Van der Stigchel 2013); or its complement, negative priming: People avoid distractor items that resemble previous distractor items (Kristjánsson & Driver 2008). In addition, there are many object-based attention effects in visual search. For example, we tend to shift our attention and gaze within, compared to between, objects (Egley et al. 1994; Theeuwes et al. 2010); and, if an attended object moves, the focus of attention follows (Theeuwes et al. 2013). We could list even more object-based effects, but our main point is: Items matter, whether we know how to define them or not. Therefore, by denying a role for items in visual search, H&O ignore, or at least downplay the importance of, a substantial part of the visual-search literature.

But how can there ever be a role for items in models of visual search if we do not even know what "item" means? Possibly, our language simply lacks the vocabulary to define "item" or "object." Many researchers, such as David Marr, have speculated that it is impossible to define "object" (Marr 1982) – and we agree. But rather than abandon items altogether (and admit defeat!) we

should adopt recent computational approaches to object recognition as an alternative to formal definitions.

Consider a modern deep-learning network: an artificial neural network that consists of many nodes across many layers. (We will not discuss one specific network, but focus on the general architecture that is shared by most networks.) Such models are inspired by the architecture of our visual system by implementing a complex arrangement of nodes, each of which only looks at small portions of the input image (Krizhevsky et al. 2012). First, this network is trained on a large set of example images, which can be either labeled (e.g., Krizhevsky et al. 2012), unlabeled (e.g., Le et al. 2012), or a mix (LeCun et al. 2010). Crucially, in all cases training occurs by example, without explicit definitions. Next, when the trained network is presented with an image, nodes in the lowest layers respond to simple features, such as edges and specific orientations (Lee et al. 2009), reminiscent of neurons in lower layers of the visual cortex (Hubel & Wiesel 1959). Nodes in higher layers of the network respond to progressively more complex features, until, near the top layers of the network, nodes have become highly selective object detectors; for example, a node may respond selectively to faces, cats, human body parts, cars, and so forth. (Le et al. 2012). These nodes are reminiscent of neurons in the temporal cortex, which also respond selectively to object categories such as faces or hands (Desimone et al. 1984, 1994). Importantly, deep-learning networks detect objects in those real-world scenes that H&O consider problematic (He et al. 2015; Krizhevsky et al. 2012); and they do so without explicit definitions, seemingly like humans do.

Combining deep-learning networks with traditional visual search models could explain how people explore their environment, item by item. As a starting point, we could take the model of H&O, and replace their bag of items with active nodes in high layers of a deep-learning network—that is, nodes that respond selectively to high-level features of the input (for example, cats), and for which the activation exceeds a certain threshold (Le et al. 2012). This would provide H&O's model with a bag of items to search through, without being fed any definition of "item." Of course, in its simplest form, this combined model is far from perfect. First, it does not explain object-based effects of the kind that we discussed above. Second, it assumes that the entire visual field is parsed at once, and does not take into account eye movements—the very idea that H&O rightfully want to get away from. But this simple combined model would be a good starting point that combines cognitive psychology with computer vision. And when combining principles from both disciplines, improvements readily come to mind. For example, a deep-learning network could be fed with eye-centered visual input that takes into account the functional viewing window.

In conclusion, we feel that H&O have been too quick to admit defeat. They have constructed a parsimonious model that explains visual-search behavior well without requiring items. Now all we need to do is put the item back in.

Where the item still rules supreme: Time-based selection, enumeration, pre-attentive processing and the target template?

doi:10.1017/S0140525X16000297, e160

Derrick G. Watson

Department of Psychology, University of Warwick, Coventry, CV4 7AL, United Kingdom.

d.g.watson@warwick.ac.uk

Abstract: I propose that there remains a central role for the item (or its equivalent) in a wider range of search and search-related tasks/functions than might be conveyed by the article. I consider the functional relationship between the framework and some aspects of previous theories, and suggest some challenges that the new framework might encounter.

Hulleman & Olivers (H&O) make a convincing case that researchers have tended to study and model search either solely from a covert attention or solely from an eye movement (EM) perspective and that if the field is to move forward there needs to be a concerted effort to combine the two—a sentiment with which I agree fully. The message is that we should replace the idea of the item with a combination of EMs and the extraction of information from fixations via a Functional Viewing Field (FVF) mechanism/perspective. EMs guide the FVF sequentially to regions from which information is extracted in parallel until the target is found. Because the size of the FVF changes as a function of target discriminability there is no role for the "item" within this framework. H&O argue that even when the task is to locate a target, the search process itself need not be item-based. Nonetheless, this of course still leaves (some) room for the item in visual search (it is the product of the search, and the target "template" will likely always be item-based).

In response, I will argue that item representations do play a central role in at least some search tasks. The "preview benefit" (Watson & Humphreys 1997) is just one finding that supports this view. In preview search, one set of distractors is presented (previewed) before a second set that contains the target. We find that people can ignore the previewed items and restrict their search to the second set of stimuli. According to the inhibitory visual marking account, this is achieved with stationary stimuli by developing a template of the locations of the old items and applying inhibition to those locations. This biases attention (and eye movements) away from those items, creating a search advantage for newly arriving stimuli. Granted, the localization of the initial items might not need to proceed via an item-by-item process (see above). However, because the inhibitory template is item- (location) based, and influences the subsequent search process, I would suggest that here "the item" (and its location) continues to play a crucial role in the subsequent search process itself. Indeed, if the locations of the old items change when the new items arrive, the preview benefit disappears (e.g., Zupan et al. 2015). In contrast, when preview items move, inhibition is applied mostly to feature maps (Andrews et al. 2011; Watson & Humphreys 1998), removing the need to track, localize, or process individual items (an example of part of a search theory in which the item is explicitly not important).

A second example in which the item probably remains salient can be found in enumeration tasks. Here people do not search for a single target but have to search for all targets (with or without the annoyance of distractors; Trick & Pylyshyn 1994) and report how many are present. In contrast to absent/present search, it is essential that items are not revisited because re-counting an item will lead to an error. With relatively coarse FVFs and an overlapping sequence of FVFs, ensuring that items are not recounted could be difficult. Perhaps here FVFs would be so small that search would effectively be item-by-item. Indeed, beyond four items enumeration appears to be especially reliant on EMs (Simon & Vaishnavi 1996; Watson et al. 2007).

Selection in time and counting things are two conditions in which the item might remain central to the task, but there are others. I wonder, for example, how contextual cuing (Chun 2000) will work without the spatial configuration of "items."

Moving on, does the FVF implicitly maintain the notion of an item? H&O argue that theories such as Attentional Engagement Theory (AET) are item-based because individual stimuli are grouped and rejected until the target is found. However, the FVF argument proposes that a stimulus emerges from the FVF which presumably is the result of some kind of competition between visual entities within the FVF. Is it possible that one episode of FVF processing equates to an entire search process in AET? So have we simply replaced the "item" from AET with more abstract visual entities within the FVF? Presumably there needs to be some individuation of "things" within the FVF for a target to emerge—aren't these "things" still just items? Notably, even though just a proof of concept, the entities fed into H&O's

simulation are discrete 1s and 0s. Are we really just arguing about how we define an item? Have we just replaced competition between items with competition between more abstract entities within the FVF?

H&O rightly state that the majority of studies have focused on relatively efficient search and this is perhaps because of the preoccupation with using small display sizes and easily separable stimuli. However, I would suggest that a focus on eye-movement-based measures could also bring with it disadvantages. For example, Watson et al. (2005) have argued that tasks that require eye movements can obscure interesting covert attentional differences because eye movements are relatively slow and noisy. In their case, the need to make EMs in a task appeared to wash out/obscure age-related attentional differences. Thus, a focus on EM measures, or worse, an encouragement to design studies that force EMs to be made, might lead to interesting effects being missed.

Finally, if we abandon the notion of the item, then what should we use to evaluate search? Will we rely just on EM frequency, or will we estimate the size of the FVF, and if so how? Do we run the risk of the circularity that H&O warn against: a search is difficult because it produces a small FVF; a small FVF is needed because target discrimination is difficult? Rather than abandoning the idea of an item altogether, perhaps we need a better way of defining what an item is.

“I am not dead yet!” – The Item responds to Hulleman & Olivers

doi:10.1017/S0140525X16000303, e161

Jeremy M. Wolfe

Departments of Ophthalmology and Radiology, Harvard Medical School, Visual Attention Lab, Brigham and Women's Hospital, Cambridge, MA 02139-4170.

jwolfe@partners.org

<http://search.bwh.harvard.edu/new/index.html>

Abstract: The item can only be dispensed within artificial tasks that, although useful in the lab, do not reflect the real world. There, the attended item is the goal of search. Hulleman & Olivers' (H&O's) model can ignore the item only by reducing search to the question of whether a patch of 0s (distractors) contains a 1 (target).

Hulleman & Olivers (H&O) describe several tricky issues in visual search, relating to relationships among covert attention, eye movements, and the inhomogeneity of the visual field. However, I believe that their proffered resolution of those issues is fundamentally incorrect. Their title promises “the impending demise of the item.” “The focus on the item as the core unit of visual search is rather problematic,” they say (sect. 4). H&O's basic misstep is to look for the core purposes of search in some tricky aspects of the search literature rather than asking why people and animals search. Outside of the lab, we are almost always searching for *something*. The goal of search may be ambiguously defined (e.g., “threat” in airport luggage or a suspicious mass in a mammogram) and that goal may be hard to segment from the background. Still, that goal is some *thing*. It is an item that we need to search for because our capacity is limited and we cannot fully process the entire scene at once (Tsotsos 1990). Following Treisman (1996), I would argue that we search because we need to attend to an item to successfully “bind” its features, and we generally need to bind features to recognize items that are the goals of search.

It is true that binding is not always necessary to do laboratory search tasks (see sect. 4.1). In some cases, the unbound image statistics are enough to classify displays as target-present or -absent. Very rapid decisions about the presence or absence of an animal in a scene (Li et al. 2002) would be one example. However,

when H&O invoke these abilities and conclude that a model, like our Guided Search (GS) model, “overestimates the role of individual item locations” (sect. 4.2), they are, again, thinking more about the lab than about the world. It may be that some “present/absent decisions are based on parallel extraction of properties of *groups* of items within local areas” (sect. 4.2), but, as we found while studying the ability to detect breast cancer in a flash, these “gist” signals are often not adequate to locate the target (Evans et al. 2013b). In real-world search, you want to know the actual location of your keys, not just that the image statistics indicate their presence at above chance levels – interesting as that may be.

H&O's model is built on the functional viewing field (FVF) that surrounds each fixation. They point out, quite fairly, that classic models such as GS have tended to ignore the inhomogeneity of the retina and the role of eye movements. As a matter of convenience, we have often used large, vivid stimuli that can be resolved anywhere in the display. We have been interested in the covert deployments of attention and have designed experiments that make the overt deployments of the eyes less critical. H&O correctly argue that models such as GS have been guilty of oversimplification in regarding eye movements as nothing more than coarse indicators of the more rapid deployments of attention. Covert attention is processing items at a rate of, perhaps, 20–40 items per second. That means that approximately 5–10 items are processed on each fixation. GS has not concerned itself very much with how those fixations are chosen; that is an omission. As H&O review, GS proposes a “car wash” model in which the 5–10 items are selected, one after the other, during fixation. As in a car wash, though they enter in series, multiple items are processed at the same time because it probably takes at least 200–300 ms to process a selected item to recognition. H&O might have been proposing that, on each fixation, all items in the FVF enter the car wash at the same time. Discriminating such parallel selection of a clump of items from rapid serial selection of each of those items is extremely difficult. Theoreticians from one camp can almost always account for the data from the other.

However, H&O are not proposing parallel selection of clumps of N items. They want to get rid of the items and parallel process everything within the FVF. This raises questions that are left for future work in their model. Consider a classic conjunction search for red vertical lines among red horizontal and green vertical distractors. How do we tell the difference between an FVF that contains a red, vertical item and one that contains red and vertical features that are not bound to the same item? It is not adequate to propose that the system can determine when red and vertical occur in the same place. Think of black vertical lines on a red background. Red and vertical are in the same location, but observers are not confused into thinking that these are red vertical targets (Wolfe & Bennett 1997).

If H&O are wrong to condemn the poor item to death, why does their model work so well? Search efficiency, as indexed by the slope of RT × set size functions, is a continuum. In the H&O model, the FVF is a parameter that scales efficiency. Hard tasks produce small FVFs and easy tasks produce big ones (Nakayama 1990). What determines FVF size? That is not clear in this schematic model so, really, FVF just serves as a free parameter to scale those slopes appropriately. The model also avoids other difficulties by coding all distractors as “0” and the targets as “1.” A parallel process, operating over the whole FVF, won't have much trouble detecting that target, but applying this model to real stimuli (e.g., conjunctions) might be a challenge.

Models such as GS also have parametric variations in difficulty that will scale search efficiency. In GS, more guidance by basic features (and by scene structure in current versions of GS) allows attention to be more efficiently directed to likely candidate targets. If you can exclude half of the items because they are, for example, the wrong color, your slope is cut in half. I strongly suspect that any mechanism for controlling FVF size will look a lot like GS's guidance, and I strongly suspect that the item, hard as it is to define, will be there to be selected.

The “item” as a window into how prior knowledge guides visual search

doi:10.1017/S0140525X16000315, e162

Rachel Wu^a and Jiaying Zhao^b

^aDepartment of Psychology, University of California, Riverside, Riverside, CA 92521; ^bDepartment of Psychology and Institute for Resources, Environment and Sustainability, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

rachel.wu@ucr.edu jiayingz@psych.ubc.ca

Abstract: We challenge the central idea proposed in Hulleman & Olivers (H&O) by arguing that the “item” is still useful for understanding visual search and for developing new theoretical frameworks. The “item” is a flexible unit that represents not only an individual object, but also a bundle of objects that are grouped based on prior knowledge. Uncovering how the “item” is represented based on prior knowledge is essential for advancing theories of visual search.

Hulleman & Olivers (H&O) present an elegant framework that aims to help us better understand visual search mechanisms. This framework proposes using fixations, rather than individual items, as the conceptual unit of visual search. The general ideas in the framework are very useful because it can account for many extant findings and identifies some shortcomings (such as embodied visual search) in the existing visual search literature.

Although this framework has its strengths, we disagree with the main argument that the item is no longer useful for understanding visual search. We do, however, agree with Olivers’ earlier argument (Olivers et al. 2011) that visual search relies on an attentional template – a prioritized working memory representation – that is typically determined before starting a task via prior knowledge and/or explicit instructions. This attentional template evolves in various ways on the shorter time scale as the task progresses (e.g., Nako et al. 2015) and on the longer time scale as the learner gains more experience (e.g., Wu et al. 2015).

We argue that the “item” is still useful for understanding visual search and developing new theoretical frameworks. Critical to our argument is the idea that the “item” (contained in the attentional template) is a flexible unit that can represent not only an individual feature or object, but also a bundle of features or objects that are grouped based on prior knowledge. Such grouping, via either explicit or implicit cues, can result in the unitization of features or objects into an “item,” which increases the amount of information held in working memory during visual search, and thus typically facilitates search performance. However, because many visual search studies control for prior experiences by using simple visual stimuli or equating prior knowledge across conditions, the nature and the limits of the attentional template are unclear. The use of prior knowledge is only mentioned briefly in H&O, but we believe that incorporating prior knowledge into visual search frameworks is critical for advancing the research area.

A growing number of studies on visual search (as well as visual working memory) demonstrate the benefits of prior knowledge on the outcomes of search tasks. For example, Nako et al. (2014a) confirmed that searching for one item (e.g., a letter) is more efficient than searching for two or more items (e.g., multiple letters), as evidenced by both neural measures (attenuated N2pc) and behavioral measures (slower reaction time and lower accuracy). Importantly, they demonstrated that if category knowledge can be applied during visual search, then one-item search and multiple-item search show very similar neural and behavioral outcomes. Nako et al. (2014b) and Wu et al. (2015) replicated and extended this initial finding using real-world objects, such as clothing, kitchen items, and human faces. In addition to prior knowledge about object category, grouping cues can also improve visual search. For example, Wu et al. (2016) showed that a heterogeneous set of novel alien stimuli grouped by an abstract rule (same versus different) can facilitate search performance.

Grouping of objects can occur not only by means of shared features and spatial proximity, but also by reliable co-occurrences

over space and time. The visual system is remarkably efficient at detecting probabilities of co-occurrences among individual objects (e.g., Fiser & Aslin 2001; Turk-Browne et al. 2005), and this ability is present in early infancy (Fiser & Aslin 2002; Kirkham et al. 2002; Saffran et al. 1996; Wu et al. 2011). A direct consequence of learning the co-occurrences between objects is that the individual objects are implicitly represented as one unit (Mole & Zhao 2016; Schapiro et al. 2012; Wu et al. 2011; Wu et al. 2013; Zhao & Yu 2016). Such unitized representations implicitly and spontaneously draw attention to the co-occurring objects during visual search (Wu et al. 2013; Yu & Zhao 2015; Zhao & Luo 2014; Zhao et al. 2013), interferes with global processing of the visual array (Hall et al. 2015; Zhao et al. 2011), and increases the capacity of visual working memory (Brady et al. 2009; see also Brady et al. 2011). These findings support the idea that individual objects can be grouped into one “item” based on prior knowledge of co-occurrences, and such representations determine the allocation of attention, group objects into chunks, and facilitate search performance.

Besides the benefits of prior knowledge on visual search outcomes, there are also costs. When asked to search for one item in a category (e.g., the letter “A”) and a foil item from the category appeared (e.g., the letter “R”), participants exhibited attentional capture to the foil at both neural and behavioral levels (Nako et al. 2014a). Wu et al. (2017) suggests that the “foil effect” is predicted by level of prior experience (e.g., distinguishing healthy and unhealthy foods based on dieting experience). Taken together, these recent studies show how the application of categorically based attentional templates (i.e., prior knowledge) can help overcome efficiency limitations in visual search by expanding the scope of target search, yet at the cost of false alarms to non-targets that fall within the search category.

In sum, we agree that investigating individual objects only may not provide a deeper understanding of visual search, but the “item” is still very useful. A better understanding of the bidirectional interactions of attention and learning allows us to build ecologically valid models reflecting cascading effects during visual search to advance this research area. Moreover, understanding how prior knowledge affects visual search and related attentional abilities has important implications for attention training. Given the growing literature showing the impact of knowledge on attention, increasing attentional abilities may involve training knowledge, rather than training attention per se.

Authors’ Response

On the brink: The demise of the item in visual search moves closer

doi:10.1017/S0140525X16000364, e163

Johan Hulleman^a and Christian N. L. Olivers^b

^aDivision of Neuroscience and Experimental Psychology, School of Biological Sciences, The University of Manchester, Manchester M13 9PL, United Kingdom; ^bDepartment of Experimental and Applied Psychology, Institute for Brain & Behaviour Amsterdam, VU University, NL-1081 BT Amsterdam, The Netherlands.

johan.hulleman@manchester.ac.uk

<https://www.research.manchester.ac.uk/portal/johan.hulleman.html>
c.n.l.olivers@vu.nl

<http://www.vupsy.nl/staff-members/christian-olivers/>

Abstract: We proposed to abandon the item as conceptual unit in visual search and adopt a fixation-based framework instead. We treat various themes raised by our commentators, including the nature of the Functional Visual Field and existing similar ideas,

alongside the importance of items, covert attention, and top-down/contextual influences. We reflect on the current state of, and future directions for, visual search.

R1. Introduction

We are grateful to all commentators for their excellent contributions. As befits a good scientific discussion, some argued that we are fundamentally wrong or went too far, while others argued that we are on the right track but did not go far enough. Again others brought to the forefront relevant aspects that we were not aware of, or had not thought of within the current context. In all cases, the commentators' views either forced or enabled us to improve our arguments and widen our perspective.

In the target article we made a number of claims, namely that (1) the study of visual search has been governed by the assumption that search proceeds on the basis of individual items; (2) this has prompted a lopsided empirical focus on easy search tasks that have relatively shallow slopes for their RT \times set size functions; and (3) this, as has been noted by others, has resulted in ignoring the eye as a major component of search. We argued that the emphasis on item-based processing has led to central, cognitive explanations of visual search (involving bottlenecks in feature binding and template matching). Yet the more likely determinant of visual search performance may be the sensory limitations associated with peripheral vision. Summarized in what we referred to as the Functional Visual Field (FVF), these limitations involve reduced acuity, increased crowding, and overall decreased attention to, and awareness of, peripheral stimuli. Together, they severely reduce the value of "item" as a concept in search. We argued that the field best abandons the item as the major unit of processing, and instead adopts the information patterns available within eye fixations as the major determinant of RTs, search slopes and RT distributions. A simple simulation showed the viability of this approach.

We have organized our response according to a number of common themes emerging from the commentaries. The first theme, treated in section R2, revolves around the question of whether we have anything new to offer. The second theme (section R3) involves the repeated objection that items, or objects, are important in visual search. The third theme (section R4) concerns the role of covert attention as a selection mechanism that is independent of eye movement. The fourth theme (section R5) comprises various issues related to top-down and contextual influences on search, as these appear to be omitted from the concept of the FVF. The fifth theme, further expanded on in section R6, is how to best define the FVF. In section R7 we focus on a number of technical issues related to our simulation. Finally, we end with a reflection on where we are after incorporating the commentaries, and discuss future directions (section R8).

R2. Does this approach offer anything new?

Several commentators remark that our framework is not new. For example, **Kristjánsson, Chetverikov, & Brinkhuis** (**Kristjánsson et al.**) state that concerns about traditional approaches are part and parcel of parallel models of visual search, and **Itti** points out that the use of fixations as

the conceptual unit is central to most computational theories and models of attention because they use prediction of human fixation locations as the main test for their efficacy. **Müller, Liesefeld, Moran, & Usher** (**Müller et al.**) note that SERR (**Humphreys & Müller 1993**) anticipated several of the ideas we advocate, and **Rosenholtz** sees clear parallels with recent work in her own lab (**Rosenholtz et al. 2012a; Zhang et al. 2015**). **Kieras & Hornof** argue that EPIC (**Meyer & Kieras 1997**) contains visual modules that already incorporate parts of the FVF concept. The FVF is also similar to what has been referred to as the "area of control" in the work of **Prinz** (**Prinz 1977**, which in turn is similar to the perceptual span in reading, **McConkie & Rayner 1975**). **Cave** suggests that Treisman and Gormican (**1988**) might have had a concept similar to the FVF in mind when they referred to the role of the resolution of the attentional scan in making local feature information available.

As we tried to make clear in our target article, we agree that many if not all components of our framework have been proposed before. A number of authors deserve a prominent position here, as they made earlier proposals about the combination of components.¹ First of all, there is **Engel (1977)**, who showed that the size of the FVF (as measured through target discrimination accuracy at various eccentricities) predicts spontaneous eye movements during visual search. Separate manual RT-related measures of search were not reported, though. **Kraiss and Knäuper (1982)** were probably the first to propose a model in which FVF size predicts number of fixations as well as manual RTs. **Geisler and Chou (1995)** also related peripheral discrimination accuracy for targets at a known location (referred to as the "accuracy window") to overall manual search RTs for the same targets at unknown locations. Although **Geisler and Chou (1995)** did not measure eye movements, they nevertheless demonstrated that a simple model assuming a variably sized fixation region (operating within the limits set by the accuracy window) accurately captured the pattern of search RTs for one of their participants. Our framework shares a number of properties with the **Geisler and Chou** model, including the idea that fixation duration is best assumed to be relatively constant, at around 250 ms. **Geisler and Chou (1995)** did not assess RT slopes or RT distributions, but here we have demonstrated that an FVF-type approach captures these across the full range of search difficulties.

Zelinsky and Sheinberg (1995) measured eye movements during search and demonstrated that manual RTs on individual trials can be accurately predicted from the number of fixations that people make on those trials, for both easy and moderately difficult searches (see **Williams et al. 1997** for similar data). Moreover, **Zelinsky and Sheinberg (1995)** proposed that it is mainly the number of items that can be processed within a fixation which determines search efficiency – an idea to which our current implementation of the FVF directly corresponds. Finally, **Findlay and Gilchrist (2003)**, in their book on active vision, also explicitly stated that by taking the limitations of peripheral vision into account we could do away with central processes of covert attention as the major delimiter in search – a view that is very reminiscent of the idea that individual items are not the unit of processing. Instead, the major delimiter is how many items can be processed in parallel within a fixation. Thus, **Findlay and Gilchrist (2003; see also Eckstein**

2011) arguably provide the most comprehensive general framework for the relationship between visual search performance and eye movements. However, they did not offer a formal model or simulation to demonstrate its viability.

The above shows that the concept of the FVF in relation to eye movements and visual search performance has surfaced at least once every decade. However, as we also alluded to in our target article, whereas the FVF has gained a clear foothold in the eye movement strand of the visual search literature, it has failed to do so in the more classic, and therefore perhaps more mainstream, strand of the literature focused on RTs and their slopes (including our own work). Our central aim was to show that this appears to be caused by the latter type of literature being heavily grounded in item-based thinking and the emphasis on covert, cognitive limitations that comes with it. This results in a fundamental clash of two classes of theory, requiring a principled choice. This choice, in our view, falls overwhelmingly in favour of the fixation-based approach. We hope our paper will aid in reconnecting all of the dots, as we believe FVF-based models have not made the impact on the field of visual search that they deserve. Furthermore, we extend the explanatory power of such models by showing that they explain search efficiency across the full range of search types, from easy parallel to extremely difficult serial, while also accounting for RT distributions and, to a great extent, errors. We believe that our effort opens up new research avenues, something that we will return to in section R8.

R3. But items are important, even essential!

Several authors indicate that we too easily abandon the item as the unit of selection. The argument comes in different varieties.

R3.1. Your data can be accommodated by item-based models

Müller et al. note that much of what we know about guidance in visual search stems from item-based experiments and that including any form of guidance for eye movements into our framework would bring it closer into line with traditional item-based models such as Guided Search. Indeed, Moran, Liesefeld, Usher, & Müller (Moran et al.), from the same lab, show that their item-based *Competitive Guided Search* model (CGS) can simulate our data, with even better fits for reaction times and errors than our own simulation. Not only does CGS show the inversion of the standard deviations from medium to hard search that we highlighted, but also its target-absent slopes for difficult search are much closer to the observed data of Young and Hulleman (2013). In addition, Moran et al. argue that our framework does not cope well with data of Liesefeld et al. (2016) who reported on a type of search where target-present search slopes are flat, but target-absent search slopes are not, and where there are large effects (>100 ms) on the intercept. Parameter adjustment in CGS allows the data to be simulated correctly. Accordingly, Moran et al. feel that it is premature to prefer a fixation-based approach over an item-based approach.

We disagree with this position for several reasons. First, we note that Moran et al. offer no accompanying graphs for fixations. We suspect this is because an item-based approach such as CGS does not allow for a principled way of including eye movements, because the rate at which items are identified in the model (θ/δ) has no fundamental relation to fixation duration. Admittedly, equating identification rate with fixation duration establishes a connection and even yields a good fit for medium search (see the final entry of Table 1 in Moran et al.). The good fit, however, comes at a price: the value of the guidance parameter nearly doubles, the quit-weight parameter increases almost a hundredfold and the residual time parameter is halved. This brings us to a more fundamental point: In our view, models are meant to engender understanding, rather than merely provide a good fit. If we compare the entries across Table 1 in Moran et al., not a single parameter remains constant. Nor are any of the parameters the same as in the paper first introducing CGS (cf. Table 2 in Moran et al. 2013; except the parameter for motor errors). What seems to be lacking is a clear justification of the parameter values used, either from experimental observations or otherwise. At the moment, CGS looks like it is too focused on fitting rather than understanding visual search data.

Furthermore, it stands to reason that a model with eight free parameters will outperform a model with only a single free parameter. But what our framework loses in lack of fit, it more than makes up for in range of description. We think that it is paramount that models of visual search explain fixations and reaction times simultaneously, rather than trying to optimise RT-modelling before adding fixations.

In this respect we see the approach of Khani & Ordikhani-Seyedlar as more promising. While they propose an item-based approach based on FIT, they do encompass fixations, which we think is crucial. In their account the first fixation produces incomplete feature maps. Rather than complete conjunctions, these maps only contain “loose” conjunctions and clusters of feature similarity, with salient features having a higher chance of entering the maps. These maps are subsequently used to covertly or overtly guide attention. Each fixation, then, leads to more detailed maps. When one or more items reach a threshold of similarity with a target template, these individual items are serially selected to establish whether one of them is the target. The number of items involved in each fixation is determined by the perceptual load (Lavie 1995; Lavie & Tsai 1994).

In a way, the proposal of Khani & Ordikhani-Seyedlar represents a return to the origins of FIT (Treisman & Gelade 1980), when the role of eye movements in search was still acknowledged. Given that it is in its very early stages, we will have to wait to see whether this proposal bears fruit, but we would nevertheless like to make two remarks. First, items actually seem to make their appearance quite late in this proposal. Only when the feature maps are detailed enough are individual items selected (cf. Hochstein & Ahissar 2002). Second, it remains to be seen how this model handles situations in which items move around. Item motion will impinge on the building up of the feature maps over several fixations, because the sources of the features – the items – will have moved position in the meantime. Yet, as shown in Hulleman (2009; 2010), search for a T amongst Ls is robust against motion.

R3.2. The importance of objects

The second type of argument in favour of items is that, clearly, we perceive discrete objects and that, equally clearly, there are object-based effects on attention (**Cave, Van der Stigchel & Mathôt, Urale**). Likewise, the goal of many, if not most, searches is to select a particular object for identification, inspection, clicking, counting, or picking up (**Eimer, Kieras & Hornof, Pasqualotto, Watson, Wolfe**). Furthermore, Watson and **Wu & Zhao** make the case that such goals are often represented through an item-based target template (but see **Prinz**, who argues against such templates and sees target detection as a disruption of integrated distractor processing across fixations). The importance of objects in visual search is clearly stated by Eimer: “[a]t a more fundamental level, it is difficult to see how objects can be replaced as conceptual units in visual search, given that the visual world is made up of objects, and finding a particular target object is the goal of a typical search task.” As Wolfe expresses it, the goal of search is some *thing*.

Yes, we wholeheartedly agree: The goal of most searches is to find a specific object. This includes most real-world searches, but also many artificial laboratory tasks, such as the compound search task in which the participant needs to make an additional decision about the target object. However, the *goal* of the process is not the same as the process itself. As **Rosenholtz** comments, “just because we recognize ‘things’ at the output of perception, and employ high-level reasoning about objects, does not mean that our visual systems operate upon presegmented things. This is a common and tempting cognitive error, which can hamper uncovering the true mechanisms of vision.” **Enns & Watson** also cite an interesting remark by Hochberg in this respect: “unlike objects themselves, our perception of objects is not everywhere dense” (Hochberg 1982, p. 214). In fact, given the severe limitations of the periphery, one could argue that real object segmentation is limited to central vision, with the periphery only being able to deliver the coarsest candidates, especially in complex, real-world scenes. It is true, as **Van der Stigchel & Mathôt** state, that recent deep-learning networks have demonstrated successful parsing of complex, real-world scenes into relevant objects. However, we note as well that, unlike the brain, such algorithms almost invariably work with images of high and homogeneous resolution.

There is indeed also clear behavioural evidence for object-based attention (e.g., Duncan 1984; Egly et al. 1994; Theeuwes et al. 2010; 2013). But although object-based attention may have a modulating influence on selection, again this does not mean that objects form the unit of selection in visual search. It may be telling that none of the referred-to demonstrations of object-based attention used visual search tasks. Rather, they typically involved rather sparse displays of at most two objects. That said, we agree with **Urale** that it is interesting to investigate how object-based attention contributes to shaping the FVF by grouping elements together—something that is also echoed by **Kristjánsson et al.** and by **Wu & Zhao**, who argue that objects may be flexibly defined by learning conglomerates of features. In fact, a fixation-based approach may accommodate such grouping mechanisms more naturally than an individual item approach (cf., Duncan & Humphreys 1989). For example, **Töllner & Rangelov**

note that increasing the number of distractors in a present/absent version of a pop-out task does not influence reaction times (although see Wolfe 1998b). Yet the same increase in a compound task benefits search. Within a fixation-based framework this can be explained by the fact that the compound task, unlike the present/absent task, requires precise saccadic targeting, which is likely to benefit more from the improved signal-to-noise ratio allowed by the grouping of the distractors. It is likewise true that there is considerable EEG evidence for the selection of individual items—some of which we will treat in more detail in section R4. Here we wish to make two comments regarding this evidence.

First, the selection of items is typically linked to the N2pc component, which is a spatially lateralized evoked potential. It does not index item selection as such, but rather the spatially selective processing of information at a relevant or interesting location. Spatial selectivity is not the same as item selectivity. Note that this type of experiment often uses sparse displays with two or four clearly segmented items (e.g., Eimer & Grubert 2014; Grubert & Eimer 2015; Woodman & Luck 1999; 2003). This makes it tempting to link the N2pc to individual item processing, but that does not necessarily follow from the evidence so far. As **Rosenholtz** suggests, we find item-based processing because we design item-based displays.

Second, the observation that spatially selective processing is stronger or takes longer for compound search tasks than for present/absent tasks, resulting in more pronounced N2pcs—as pointed out by **Töllner & Rangelov**—might simply reflect the fact that compound tasks require more fine-grained discrimination, not that this discrimination is item-based. While these types of EEG experiments are extremely useful in uncovering attentional processes, in our view they do not provide direct evidence that visual search is item-based.

We emphasize that we do not claim that visual search is *never* object- or item-based (as **Van der Stigchel & Mathôt** appear to suggest), nor are we of the opinion that we should not use artificial displays consisting of clearly separated items. As we pointed out in the target article, in displays with separate items where target-distractor discrimination is very difficult, inspection of each individual item may be required (resulting in a corresponding FVF). Furthermore, in some tasks target-distractor discrimination is easier, yet still an item-based strategy is required. As **Watson** suggests, one such task is *counting* multiple items rather than finding a single one. In a counting task the identity of the individual item matters, because it is important to separate those that already have been counted from those that have not. In other words it is important not only to discriminate targets from distractors, but also to discriminate between targets. In our view this necessitates smaller FVFs to prevent interference from similar but already counted items. Support for this contention comes from the last experiment in Hulleman (2010), where participants had to establish whether there were at least five Ts in a display that also contained Ls. When there were either very few or very many Ts in the display, motion of the items did not influence performance. When there were four, five, or six Ts, however, more errors were made when the items were moving. This demonstrates the interaction between task demands and FVF size. Only when individual item identity is crucial does

the FVF become small and does search become item-based.

R3.3. The importance of feature binding

Because items are defined as bound features, item-based approaches to visual search are predicated on feature binding. **Wolfe** is the most explicit here in arguing “that we search because we need to attend to an item to successfully bind its features, and we generally need to bind features to recognize items that are the goal of search.” Wolfe agrees that binding is not always necessary but also points out that unbound features do not allow for accurate localisation of the target. As a case in point, he wonders how it would be possible to determine the presence of a red vertical target amongst red horizontal and green vertical distractors without binding the red and the vertical into a single item by attending to it. **Eimer** is less explicit but nevertheless appears to concur with this objection by noting that “[w]hat remains unclear is whether such global area-based mechanisms can detect the presence or absence of targets even in moderately difficult search tasks where no diagnostic low-level saliency signals are available and distractors share features with the target.”

In our view this type of argument is based on a couple of assumptions that do not necessarily hold.

First, consider **Eimer**'s assumption that there are no diagnostic low-level saliency signals available in moderately difficult search where distractors share features with the target. We assume that for **Eimer**, as for **Wolfe**, this means a conjunction search, for example, for a red vertical amongst red horizontal and green vertical distractors. Classic item-based thinking dictates that for these types of searches the target item does not carry a unique feature. However, this assumes that the individual features of the individual items provide the *only* type of information available and that indeed there are no other diagnostic signals possible. This neglects that at several levels a patch containing a red vertical target among red horizontal and green vertical distractors is different from the same patch without a target. Even if the items within a patch do not have distinctive features, the overall image of the patch does. For instance, in our colour/orientation conjunction example, a patch with n distractors will contain x reds, x horizontals, $(n-x)$ greens and $(n-x)$ verticals. So there is a match between red and horizontal and between green and vertical. Replacing one of the distractors with the red vertical target will, depending on the distractor replaced, yield either x reds, $(x-1)$ horizontals, $(n-x)$ greens and $(n-x+1)$ verticals or $(x+1)$ reds, x horizontals, $(n-x-1)$ greens and $(n-x)$ verticals. In both cases, there is now a mismatch between red and horizontal and between green and vertical. So without assuming any item-based binding, it is possible to distinguish between patches with and without a target purely on the basis of summed totals. It is indeed necessary to bring information from colours and orientations together but, as this example shows, this does not have to happen at the level of a fully bound individual item. **Rosenholtz** lists quite a number of other properties of target and non-target patches that apply here. The presence of a red vertical changes the local spacing and alignment between the red bars, resulting in what resembles red T-like or L-like junctions. It likely changes the spatial frequencies present in the patch,

because the red vertical target would probably be adjacent to a green vertical. There is no a priori reason to assume that these signals are not available for visual search.

Second, feature binding is only needed if one assumes that, unless attended, features remain represented separately throughout the visual system. As discussed by Di Lollo (2012) this assumption was originally based on the work of Hubel and Wiesel (1962; 1977) whose single cell recordings indicated that neurons in the primary visual cortex responded selectively to orientation or colour but not both. This then implied the need for a mechanism to integrate these features at a later stage by binding them, because a neuron representing one type of feature does not “know” about the neuron representing the other feature. However, it has long since been established that many neurons, also in early visual cortex, respond to integrated features (e.g., Friedman et al. 2003; Seymour et al. 2009; 2010; Shipp et al. 2009). Moreover, there is considerable cross-talk within and between retinotopically organized layers in both feedforward and feedback pathways. This altogether makes binding less of a problem than it was (Di Lollo 2012; Hochstein & Ahissar 2002; Lamme & Roelfsema 2000). Because many of these integrative mechanisms appear to operate without attention, the role that attention plays in binding remains unclear. We suspect that attending to items may be necessary to distinguish their feature conglomerates at a sufficiently fine resolution, rather than to bind these features together – especially when it involves overt orienting (but not exclusively so; He et al. 1996; Hochstein & Ahissar 2002).

In sum, it is questionable (1) whether binding of item features is really necessary (2) what relative contribution attention makes, and (3) whether the visual system cannot use more global features other than those of the individual items. All in all, we find an insufficient basis for the claim that binding necessitates item-based selection in visual search.

R4. What about covert deployments of attention?

Several authors raised the concern that a focus on eye movements and/or FVFs faces the dilemma of how to explain search in the absence of saccades (**Ohl & Rolfs**), because our framework must assume that there is then nothing left for covert attention to do (**Cave and Kristjánsson et al.**). This is at odds with clear evidence for covert attention effects that often precede eye movements (**Eimer, Khani & Ordikhani-Seyedlar**). Others raise the more general concern that a focus on eye movements may result in missing interesting covert attention effects (**Lleras, Cronin, Madison, Wang, & Buetti [Lleras et al.], Watson, Wolfe**). Specifically, Wolfe writes that his interest in serial covert deployments has been the reason to design experiments that make eye movements less critical. Lleras et al. state that our framework assumes that parallel searches (occurring within a single fixation) are not very interesting because they are all created equal. They point out that this misses out on subtle but important variations, due to differences in task sets and differences in similarity between the target item and distractor items, that occur even in parallel search without eye movements. This means that another source of variation in visual search stems from the efficiency with which individual items are judged in parallel search.

We would like to reiterate that we present a *fixation* based account of visual search, not an eye movement based account. This is a subtle but important difference. As we wrote, visual search can clearly occur without eye movements. Yet even when observers are instructed to keep their eyes still, at least one fixation is involved. Accordingly, search will be limited by the retinal and cortical constraints imposed by that fixation, leading to reduced discriminability in peripheral vision. This alone means that not all searches within a fixation are equal: Even when a target falls within the FVF and the mechanism is in essence a parallel one, detection rates will not be homogeneous, as the signal-to-noise ratio is worse in the periphery (e.g., Geisler & Chou 1995). Furthermore, signal-to-noise ratios will differ for different stimulus combinations and set sizes, leading to either subtle (e.g., Buetti et al. 2016) or less subtle set size effects (taken as indicative of serial search). Specific top-down task sets may further shape the priority map, boosting some signals over others, as in Guided Search.

The end result of the interaction between bottom-up and top-down factors is likely to be a covert selection of a candidate region for further evaluation (e.g., on a secondary detail) or motor response (e.g., an eye movement, see also Cave's commentary). This region might be an item, but that is not necessary. In natural circumstances covert selection is followed by an eye movement, which enables a full resolution image of the target. In some tasks this is even explicitly required (e.g., Buschman & Miller 2009). In other tasks eye movements are forbidden, but this does not necessarily stop the underlying selection process. Furthermore, the process is not perfect, so multiple covert hot spots may occur when a selected region turns out to be the wrong one. This is especially the case when non-targets are deliberately designed to closely resemble targets. For example, Woodman and Luck (2003) measured EEG when participants saw visual search displays with two salient target candidates (both defined by colour against black distractors), while maintaining fixation. They showed that the target candidate near the fovea was selected first, followed by selection of the more peripheral one (as the central one turned out not to be the target after all). In another experiment (Woodman & Luck 1999) they showed that observers first (again covertly) selected the target candidate that carried the most frequent target colour, followed by the candidate that carried a less likely colour (see also the commentary of Töllner & Rangelov for target prevalence effects on the EEG signal). Both findings are perfectly consistent with a non-homogeneous FVF, where essentially parallel bottom-up and top-down processes deliver target candidates, but not necessarily all at the same time.

One could argue that in these specific EEG cases seriality is strongly imposed by the task, given that it is more efficient to inspect the closest or most prevalent item first. As recent work by Eimer and colleagues has shown, there is virtually parallel processing of multiple target candidates when the task is more balanced (Eimer & Grubert 2014; Grubert & Eimer 2015). Note that this is essentially no different from signal detection accounts of search that also assume parallel processing (Eckstein et al. 2000; Palmer et al. 2000; Verghese 2001). Furthermore, when one forces participants to keep the eyes still, as is usually the case in EEG experiments, the relative influence of covert

selection will increase. But this is usually not the case in the real world, nor in most laboratory search tasks for that matter.

We do not deny the existence of covert attention, nor the importance of its investigation. Rather, our point is that some visual search theories appear to rely too heavily on covert selection to explain search, in that they assume that search efficiency is predominantly determined by fast covert scanning of items in combination with central, item-based bottlenecks (processing items at a rate of 20–40 items per second, as Wolfe confirms in his commentary). The fixation-based view follows Findlay and Gilchrist's (2003) stance that covert selection is more likely to be the *end product* of a search process that is primarily determined by limitations in retinal and cortical receptive fields, at most delivering the candidate region for the next fixation. As such, it is not an independent search process that occurs during fixation, but part of the active eye movement mechanisms. As Findlay and Gilchrist (2003) pointed out, there is little to no evidence for a serial covert scan during fixations. Further evidence against covert serial scanning comes from the robustness of search against motion of the items, even for searches that have been deemed serial (Hulleman 2010; Hulleman & Olivers 2014; Young & Hulleman 2013). Serial deployment of covert attention within a fixation would predict a drop in performance, because it becomes harder to distinguish between items that have and have not yet been inspected.

R5. Where to look next: top-down factors

A number of authors point out that an account solely based on an FVF is incomplete because it fails to incorporate important if not crucial mechanisms that determine where people look next in a visual scene. Here we best summarize such mechanisms as “top-down” in nature, and they come in a number of varieties: Enns & Watson; Lleras et al.; Shi, Zang, & Geyer [Shi et al.], Töllner & Rangelov, and Van der Stigchel & Mathôt emphasize the importance of the *task*, whereas Menneer, Godwin, Liversedge, Hillstrom, Benson, Reichle, & Donnelly [Menneer et al.] and Itti, as well as Crabb & Taylor; Crawford, Litchfield, & Donovan [Crawford et al.]; Laubrock, Engbert, & Cajar [Laubrock et al.], and Watson highlight the role of *context* and *scene gist* in making predictions, guiding selection, and determining scan strategies. *Learning*, whether explicit or implicit, or *task expertise* also play an important role in shaping search (Kristjánsson et al., Van der Kamp & Dicks, Wu & Zhao, Crawford et al., Menneer et al., Van der Stigchel & Mathôt). Kieras & Hornof argue that a full model of such task- and memory-dependent strategies therefore requires an overarching cognitive architecture like EPIC (Meyer & Kieras 1997) or ACT-R (Anderson & Lebiere 1998). Müller et al. note that some form of feature guidance, such as in Guided Search, is required in many search tasks, and Wolfe suspects that these guiding features will be the same as those determining the FVF. Finally, Henriksson & Hari suggest that top-down cues need not be represented at sensory levels but may be of a high-level semantic and even social nature.

Non-visual influences are also accentuated by **Van der Kamp & Dicks**, as well as **Campion** (see also **Kieras & Hornof**; **Töllner & Rangelov**), who underline the influence of *motor requirements* on the search process. Because observers normally move about in their environments, there is a continuous perception-action cycle. **Van der Kamp & Dicks** call for a move away from traditional search tasks where observers are passively watching computer screens. **Campion** even calls for a move away from the information processing view of cognition that such traditional approaches appear to induce. These views are reminiscent of the Gibsonian ecological approach. They also fit with the active vision approach of Findlay and Gilchrist (2003) that precedes our own. On the other hand, in referring to Julian Hochberg's legacy, **Enns & Watson** appear to endorse the information processing approach, stating that, "what happens behind the observer's eyes is more important than what happens in front of them (the display items) or even in them (the FVF)."

We welcome all of these important suggestions. We agree that task, context, and actions play an important role in driving selection in laboratory studies, and even more so in real world environments. We also agree that in real-world circumstances, fixation patterns are generally not random. We restricted our simulations to standard abstract laboratory displays—where such influences are minimized or controlled for—exactly because these have typically subserved the RT data that item-based theories have been grounded in. We sought to demonstrate that these types of data can be more straightforwardly captured by a model that used fixations rather than display items as its unit of processing. Given the high level of randomness of laboratory displays, a simulation that assumes random fixation selection (with some restrictions) suffices.

Nevertheless, the decision about where to look next is one of the major research questions arising from a fixation-based approach. We believe that FVF-based accounts provide a more natural and fruitful way of thinking about how this decision comes about than item-based accounts. First of all, we agree with **Wolfe** that features that have shown a high degree of guidance in classic visual search tasks will result in large FVFs. However, whereas Guided Search assumes feature status, and thus the capability of guiding attention, for at most a handful of visual properties (**Wolfe & Horowitz 2004**), within an FVF account *any visual information that can be discriminated beyond the fovea can, by definition, subserve attentional guidance in visual search*. Be it a low-level feature, a semantic category, or the social signal conveyed by a complex facial expression, if observers can distinguish it in the periphery, it can become the next target of fixation. The central point is that such information need not be item-based. What we claim is that once the attention-guiding properties are mapped out, the item as such is no longer necessary for explaining search mechanisms. Note here too **Rosenholtz's** remark that the same information available in a rich set of image statistics (**Keshvari & Rosenholtz 2016**) also plausibly underlies scene perception (**Ehinger & Rosenholtz**, in press; **Rosenholtz et al. 2012b**). This suggests a common encoding scheme for both extracting the scene context and supporting search. In this respect, Guided Search can be regarded as a representative of classic *early selection* theories, in which only relatively low-level properties can be used to filter information (**Broadbent**

1958; **Treisman & Gelade 1980**). FVF-based theories, on the other hand, are representatives of multiple level selection theories (**Allport 1980**; **Findlay & Gilchrist 2003**; **Norman & Shallice 1980; 1986**), in which the level of selection is determined by the task requirements and the level of information available in the input.

In our view then, the FVF is not simply a re-description of bottom-up salience. Classically, the FVF for a certain type of information is measured with a task in which observers actively look for this information in a known location. The FVF is thus an amalgamation of the availability of the information in the input, and the top-down modulation of that input. Indeed, it has been shown that the FVF can change size or shape depending on additional task load or the expected spatial distribution of the target information (e.g., **Engel 1971**; **Ikeda & Takeuchi 1975**; **Williams 1982**). This might at least partially aid in natural scene search, where targets are often restricted to certain spatial areas. Given the well-documented effectiveness of feature- and object-based attention, the shape or size of the FVF is also likely to be modulated by increasing the gain on specific feature or category distinctions, but to our knowledge there have been few studies looking directly into this (e.g., **Pöder 2007**, has shown how repeating the target feature reduces peripheral crowding, but did not map out the full FVF). Thus, **Enns & Watson's** assertion that "what happens *behind* the observer's eyes is more important than what happens ... *in* them (the FVF)" (our italics), is partly tautological: The mechanisms determining the FVF *include* what happens behind the eyes. That said, **Enns & Watson** are correct in suggesting that in setting up our account, we wished to emphasize the sensory restrictions in visual processing outside of the fovea, rather than the central cognitive restrictions associated with foveal processing.

Of course, the amalgamation of top-down and bottom-up factors into a single construct makes it vulnerable to becoming circular and unfalsifiable. We will address this issue in section R6.

R6. The nature of the FVF

Several authors have questions about the nature of the FVF or whether it is even possible to come up with an operational definition. **Phillips & Takeda** feel that the FVF lacks independent motivation, a sentiment also expressed by **Kristjánsson et al.**, **Watson**, and **Little, Eidels, Houpt, & Yang [Little et al.]**, who mention the risk of circularity: Search is difficult because the FVF is small, and the FVF is small because search is difficult. Furthermore, according to **Itti**, positing a single FVF size conflates guidance, selection, and enhancement mechanisms. The relation between the FVF and guidance is also touched upon by **Wolfe**, who thinks that the mechanisms controlling the size of the FVF will look a lot like those controlling guidance in Guided Search. Control of the size of the FVF also comes to the fore in the various comparisons of the FVF to a spotlight (**Laubrock et al.**; **Itti**), a zoom lens (**Cave**), an attentional window (**Kristjánsson et al.**) and its relation to perceptual load (**Khani & Ordikhani-Seyedlar**). **Rosenholtz** points out that is important to make clear that the FVF is not a *mechanism*, where its size is under active control, but an *outcome* of several mechanisms: It describes the informative visual regions for a particular task.

R6.1. Circularity

We will first address the issue of circularity in the definition of the FVF. Whereas **Phillips & Takeda** argue that it is possible to provide a mathematical basis for the FVF, we believe that an approach to visual search based on the FVF allows the circle to be broken in an empirical manner as well, because it connects performance in visual search to performance in other visual tasks. Several researchers have already led the way. For instance, Engel (1977) determined the “conspicuity area” by having observers detect the appearance of a target. He subsequently tested visual search for that target and related the conspicuity area with the cumulative probability of finding the target. A similar approach was used by Geisler and Chou (1995): They measured an “accuracy window” in a two-alternative forced-choice (2AFC) task, where one of the intervals contained the target and the other contained only distractors. They then found a correlation between the size of the accuracy window and reaction times in a visual search task. So far the number of different types of visual stimuli that have been tested and correlated with search this way has been limited, but this is a matter of filling in the gaps.

We do note that the method of Geisler and Chou (1995) could still be considered circular, because the 2AFC task used for measuring the FVF is essentially a skeleton version of the search task, and it would therefore be a surprise if there were no correlation. This criticism holds a little less for the approach used by Engel (1977), who presented single targets without distractors in the FVF task. But here too one could say that detection predicts detection. Still, we want to argue that this method brings us at least one step further than the item-based approach.

Some non-psychophysical methods might also offer a route out of circularity. One such method is the use of gaze-contingent displays. Young and Hulleman (2013) demonstrated that a very difficult search task is more robust against the masking of non-fixated items than easier search tasks. This indicates that the former has a smaller FVF than the latter. Neurophysiological data, too, may be used as an independent predictor of search performance. For example, Song et al. (2015) recently reported how anatomical characteristics of V1 and V2 cortex predict individual differences in both the precision of neural population tuning and performance on a visual discrimination task at various eccentricities. It would be exciting to see whether the same measures also predict visual search performance.

In summary, although we agree that circularity in the definition of the FVF presents a problem, we do think it is possible to find a solution that will provide a size measure of the FVF that is independent of search performance. Therefore, although **Wolfe** rightly observes that the size of the FVF acts to scale search slopes in the same way that guidance does in Guided Search, we think that there is a crucial distinction: Only the FVF account seeks to systematically anchor search performance in an independent task.

R6.2. Control over FVF size

The issue of size control goes to the very heart of the nature of the FVF. We agree with **Rosenholtz** that the FVF is an outcome – rather than a mechanism – with its size delimited by the interaction between retinal and cortical

constraints on the one hand, and task demands on the other. Sometimes the retinal and cortical constraints might work in opposite directions (for instance in the Gestalt grouping mentioned by **Urale**). In any case, the size of the FVF is not actively controlled, but a particular task can only be performed when the FVF does not exceed a certain size. Within the FVF, active modulation (for instance by attentional processes or central perceptual load, **Khani & Ordikhani-Seyedlar**) might be possible, but this can never be more than modulation. Moreover, this active modulation is not item-based.

As such, the FVF is fundamentally different from attentional zoom lenses and spotlights. The latter are explicitly conceived of as operating covertly, independent from the eye, while the FVF is centred on current fixation. As LaBerge and Brown (1986, p. 198) put it: “attentional factors dominate in processing visual targets [...] retinal sensitivity factors have a minor role, if any.” We therefore do not agree with **Cave** when he suggests that Treisman and Gormican’s (1988, p. 17) description of the role of spatial attention is similar to the FVF. Indeed, Treisman and Gormican wrote: “Attention selects a filled location within the master map and thereby temporarily restricts the activity from each feature map to the features that are linked to the selected location. The finer the grain of the scan, the more precise the localization and, as a consequence, the more accurately conjoined the features present in different maps will be.” But this alludes to the way covert attention promotes correct feature binding. It ignores the fact that the change in real spatial resolution from peripheral to central vision by overtly fixating an item probably contributes much more to correct object perception.

FVF size control is also at the core of **Itti**’s commentary. He suggests that FVF size is unlikely to be fixed as in our simulations, and that an FVF that is allowed to rapidly change size and form becomes a liability because it would be very difficult to measure in real time. But as pointed out above, we do not see the FVF as an entity that is directly under active control of the observer. Rather, it is the outcome of the interaction between task demands on the one hand and retinal and cortical limitations on the other. Task demands might change over the course of a search, but this will be in a gradual, predictable manner (and is moreover something that any model of search will have to deal with.) Any change in FVF size will follow this change in task demands. Itti also suggests that it may be necessary to separate the FVF into three: a broader FVF for guidance of search, a smaller FVF for selection, and a potentially even smaller FVF for enhancement, because he thinks that using a single FVF size conflates the separate mechanisms of attentional guidance, attentional selection, and attentional enhancement. However, conflation might not necessarily be a drawback. The advantage of using the FVF is that these three presumably separate mechanisms of increasingly “homing in” on the target might actually represent one and the same process operating on increasingly detailed and target-like information. In a recent paper, Zelinsky et al. (2013) argued that guidance and recognition in visual search are two sides of the same coin: eye movement guidance is in essence recognition from the corner of the eye. The first signal that may be recognized is some fuzzy statistical reflection of what might be a target that, after an eye movement has been made, becomes recognition of a more detailed version.

R7. Technical issues

Several authors take issue with some of the more detailed choices we have made in our simulation.

R7.1. The stopping rule

Moran et al. suggest not only that our simple stopping rule leads to poor fits for the target-absent trials in difficult search, but also that replacing it with a more plausible rule will substantially change search RT distributions and error rates, possibly to an extent that puts the desirable properties of our framework at risk. We disagree. In our view, a more plausible stopping rule will actually improve the quality of our simulation. Currently, simulations of target-absent trials are most affected by the poor quality of our simple stopping rule. Furthermore, the influence of the stopping rule increases with decreasing FVF size. In combination, this means that the difficult target-absent condition is most affected. However, this is also the condition where the discrepancy between simulation and observed data is largest. Therefore we expect improvement with a more plausible stopping rule. Crucially, this should lead to fewer target-absent trials with extremely long reaction times and reduced variability in the target-absent trials of difficult search. Consequently one of the major insights of this paper (reversal of SD from medium to difficult search) remains unscathed. So, although a more plausible stopping rule is certainly needed, implementing it is not expected to invalidate the fundamental strength of our framework.

Crawford et al. raise an issue related to the use of the stopping rule. They applied the framework to data collected from radiologists who assessed chest radiographs, and noted that there may be an incompatibility in how people reach a target present or absent decision. Crawford et al. report that in their data, target-absent decisions are faster than target-present decisions, whereas in our framework (and in about every other model of visual search based on fundamental research) target-absent trials are typically slower. Although this could point to basic differences between the fundamental and applied research domains, we believe this is a case of incommensurate definitions of “absent decision.” In our approach a target-absent reaction time is the result of searching close to an entire display, failing to find the target and terminating the trial with a target-absent response. This typically takes longer than searching the display, finding the target and terminating the trial with a target-present response. For Crawford et al. a target absent reaction time appears to be something else: A radiologist fixates a particular part of the radiograph, correctly decides that there is no suspicious lesion and moves on. The target-absent reaction time is taken as the duration of the fixation of this lesion-free zone. Target-present reaction times are defined in a similar way. Only here the radiologist correctly decides that there is a suspicious lesion. So when Crawford et al. refer to target-absent and target-present RTs, they refer to single fixation events, while in standard search tasks a decision RT is the result of the accumulation of several such decisions for the same image. Their comment highlights that the results of fundamental and applied research cannot productively inform each other without consistent definitions.

R7.2. Fixation durations are not constant

Enns & Watson, Henriksson & Hari, Laubrock et al., Little et al., Menneer et al., and Shi et al. all point out that fixation durations vary, for example, in response to task demands, and that the constant value used in our simulation is therefore unrealistic. **Ohl & Rolfs** make a similar point in stating that there is useful information in the amplitude of saccades, including microsaccades. Little et al. make the valid argument that allowing variability of fixation durations will lead to increased variability in reaction times, and that this variability is positively correlated with the number of fixations. As a result the target-absent condition of difficult search will see the largest increase in variability, perhaps even making the target-absent trials more variable than the target-present trials. We ran some additional simulations with variable fixation durations. It is indeed possible to make difficult target-absent trials more variable than target-present trials, but this takes levels of variability in fixation duration that go far beyond those observed in our own recorded eye movement data. Variable fixation durations will thus not change the basic pattern of our simulation. We also point to Zelinsky and Sheinberg (1995), as well as Geisler and Chou (1995), who showed that it is sufficient to assume a relatively constant fixation duration.

R7.3. Attentional dwell time and its relation to fixation duration

Eimer questions our equating dwell time with fixation duration. He points out the discrepancy between our fixation duration (250 ms) and attentional dwell time estimates (300–500 ms). **Menneer et al.** also suggest that there might be a dissociation between fixation location and the location whose information is currently processed when searching scenes. We would like to reply that although there seems to be a consensus that dwell time estimates of 20–40 ms per item in visual search tasks are too low, there is actually much less consensus on a more realistic value. The 300–500 ms mentioned by Eimer seems to be derived from Duncan et al. (1994). However, note that this is for tasks where two difficult-to-perceive (because masked) *targets* need to be reported. Consolidating targets for report likely involves additional processing and is quite different from the standard visual search task where the mere presence or absence of a single target is to be reported. Furthermore, most estimates of dwell time come from studies that did not use a visual search paradigm and presented items sequentially rather than simultaneously. Some of these dwell-time estimates are in the region of 250 ms (Theeuwes et al. 2004) and even go as low as about 100 ms (Wolfe et al. 2000). All in all, this suggests dwell times may be substantially lower than 500 ms.

We do agree with **Menneer et al.**’s point that there might be differences between scene search and more standard search tasks in terms of the relation between fixation duration and dwell time. This is one of the challenges facing the development of a unifying framework. One way to accommodate dwell time variations would of course be to consider them as contributing factors to the fixation duration variations mentioned in section R7.2.

R8. Conclusions: Where do we stand and where do we go?

In our view one of the most significant outcomes of this discussion is that all commentators seem to agree that it is important to include fixations in theories of visual search. This would constitute a major change for the item-based, attentional strand of the visual search literature, which should have wider implications. Referring to their own 2003 work, Findlay and Gilchrist (2005, p. 259) wrote: “There has indeed been widening interest more generally in eye scanning and we have even been prepared to suggest that a fundamental theoretical shift is in the process of occurring.” Yet more than 10 years down the line, a simple look at some of the most popular current textbooks on cognitive psychology, cognitive neuroscience, and even perception shows that the treatment of visual search seldom includes eye fixations, let alone assigns them a central role² (Braisby & Gellatly 2012; Eysenck & Keane 2015; Goldstein 2014; 2015; Reisberg 2013; Sternberg, 2017; Ward 2015; Wolfe et al. 2015). In fact, most do not go further than Treisman’s FIT plus perhaps some excursion to Duncan and Humphreys’ AET or Wolfe’s Guided Search. Not only do our students grow up with these textbooks, but also the books serve as a theoretical frame of reference for many researchers, including clinicians, who use search tasks only as a tool, rather than as a topic of study. Such researchers may not only miss out on a rich empirical source of information, but also may attribute their findings to the wrong mechanisms.

That said, it is also clear that the idea of abandoning the item as the conceptual unit of visual search does not enjoy a similar consensus. The commentators have offered three main arguments for keeping the item: (1) item-based models can accommodate our data; (2) objects are important in behaviour; and (3) feature binding is necessary. However, we do not think that these provide sufficient support for the suggestion that visual search is essentially an item-based process. Neither do we think that the effects of covert attention in visual search make this case. So although the commentators have sharpened our views, they have not changed them: We still see a fixation-based framework as the best way to think about visual search.

In essence, the fixation-based approach returns visual search from mainly being an attentional problem to mainly being a perceptual problem. By using fixations and an FVF based on retinal/cortical limitations (in combination with task demands) the proposed framework makes direct connections with processes involved in crowding, reading, and perception in general. For example, there is no a priori reason to assume that there are no FVFs for semantic or categorical information (Wu & Zhao; De Groot et al. 2016; Lupyan 2008), which after all form an intrinsic part of the perceptual process.

Apart from making a connection between visual search and other areas in vision science, we also intended our paper as an attempt to invite new thinking about current problems in visual search, whether in fundamental or in applied research. We feel encouraged that several of the commentators have already made a start with this. **Pasqualotto** demonstrates how abandoning the item-based approach to visual search facilitates thinking about the common aspects of visual and haptic search. **Kristjánsson**

et al. argue that the FVF approach makes certain testable predictions about priming, specifically that if search is fixation-based, priming ought to be fixation-based too. **Menneer et al.** also derived new predictions from our framework, which prompted them to reanalyse their data for the prevalence effect in visual search through X-ray images (where rare targets are more easily missed than frequent targets). Their findings were consistent with our framework. **Crawford et al.** note that expert radiologists examining chest radiographs have scan paths that differ from those of novices, with fewer fixations and larger saccadic amplitudes. This shows the potential of fixation-based approaches to incorporate expertise and learning. **Van der Kamp & Dicks** point out that successful goalkeepers also have fixation patterns that differ from those of less successful ones, citing Piras and Vickers’ (2011) observation that experienced goalkeepers seem particularly interested in the *empty space* between the non-kicking leg and the ball. Empty space has no role in item-based theories. An FVF approach is more flexible in accommodating this kind of observation, as it does not have to rely on the features of individual objects. There are many aspects in the latter and other real-world searches that differ from standard laboratory search (see also **Crabb & Taylor**) and that are not incorporated in our framework, such as the effects of ill-defined targets, low target prevalence, and unknown number of targets. But by adopting a common framework and common definitions, we believe it will be easier to establish what these differences actually are and what factors give rise to them.

Clearly, there is much more work to be done. First, as also became clear from the commentaries, the FVF needs to be defined properly, through extensive empirical measurements using multiple independent and converging methods, as well as through clever computational modeling (cf. **Itti**). Interestingly, in 1995 Geisler and Chou (p. 361) wrote that “the low-level mechanisms are not understood well enough at this time to precisely quantify the variations in search information across different search stimuli.” In our view, this holds more than two decades later. There has been no concerted large-scale effort to map out the FVF for the wide range of stimuli that the visual system is sensitive to. Or as **Kieras & Hornof** put it, we still need to collect the empirical data to more completely parameterize the detectability of visual properties based on object eccentricity, size, and density. One reason for not embarking on this effort may be the vast task that lies ahead: By definition, there will be a specific FVF for every type of stimulus contrast. Add to this, the effects of spatial attention, top-down attentional sets, context, and experience. Clever down-sampling is therefore required. A second reason may have been the perceived circularity. As we have argued, there are methods in place that at least partially address this.

Little is also known about the dynamics of FVF-based search. How does the visual system determine the length and precision of saccades and the duration of fixations? The FVF is an outcome, not a mechanism, but it appears that this outcome can be used to make the system adapt, perhaps based on the initial fixation or on previous experience with similar displays. The dynamics become even more complicated when we consider that the FVF changes during the search itself. This occurs when the

required information changes, for example, from the target-defining feature to the to-be-reported features (as in compound search).

The second area where work clearly needs to be done is on how the next fixation location is determined. As the commentators made clear, the roles of task, social, and scene context; learning and expertise; and action requirements are grossly underspecified. To this list we would like to add mechanisms that enable efficient sampling of the display, for example by selecting certain clusters of items, or the space in-between (e.g., Zelinsky et al. 1997; Pomplun 2007). Furthermore, any of these effects is likely to be amplified in real world situations, with more realistic task goals, expectations, and actions. However, we wish to point out that these problems are not specific to our account; item-based approaches have to explain such effects as well. Our message is that the problem of where to look next is best approached from the standpoint of the machinery that does the looking, and that the outcome will also be informative for the cases where one only considers covert attention. Note that the still-dominant theories of visual search (FIT, AET, Guided Search) were all developed at a time when eye movement recording was in its infancy, monitors and computers capable of displaying photorealistic pictures were yet to be introduced, and thus investigations into the perception of objects was confined to search tasks using simple, clearly defined, and static items. Rather than retrofitting the dominant theories to more complex tasks and scenes, we think that it is better to use a theoretical framework that starts with what all of the tasks, simple or complex, lab-based or “real-world,” have in common: fixations.

NOTES

1. We think that some of these works deserve a more prominent position than we gave them in the target article. We initially missed out on Zelinsky and Sheinberg's (1995) book chapter (which in our minds had somehow merged with the somewhat similar Zelinsky & Sheinberg 1997 paper) and Geisler and Chou (1995), as we failed to fully scan the literature for another term for the functional visual field, namely “visual lobe.” We included references to both works after acceptance of the manuscript. While processing the commentaries, we found out that we also blatantly missed Findlay and Gilchrist's (2003) *Active Vision* book. We thought that we had covered Findlay and Gilchrist's stance in treating their earlier and later work (Findlay & Gilchrist 1998; 2001; 2005), but their book is much more explicit in linking overt attention, the FVF, and manual RTs. We apologize for these and other oversights, and we would like to pay tribute to these authors here.

2. This raises the question of why the paradigmatic shift predicted by Findlay and Gilchrist (2003) has not yet happened. We can only guess, but one factor in the continued popularity of item-based theories may be the satisfying way in which they combine intuitive appeal, a clear rationale for search slope differences and the complex workings of early visual cortex.

As pointed out by several commentators, we perceive the world in terms of objects. This provides a strong intuitive foundation for item-based theories. Item-based theories also offer a relatively straightforward understanding of the fundamental result in visual search: In most searches RTs increase with display size, but in some they remain constant. At the same time, item-based theories allow for the complexity of early visual cortex, specifically the fact that features appear to be represented separately. They then elegantly explain how attending to the item allows the features to be recombined to form the items we see. Adding eye movements and their associated limitations to this account of visual search complicates it substantially.

A further factor might be that the item-based approach has been supported by strong computational models, demonstrating its explanatory power for manual response data from a range of search tasks, including pop-out tasks. Models of eye movements in visual search have typically been confined to rather difficult search tasks because they require eye movements. As such, bar the demonstration by Zelinsky and Sheinberg (1995), there has been little attempt to model search slopes and other manual response data.

We hope to have shown that a fixation-based approach to visual search can also have these two ingredients for success, by demonstrating how the implementation of a conceptually simple fixation-based framework can simulate the entire range of search slopes, from pop-out to extremely serial.

References

[The letters “a” and “r” before author's initials stand for target article and response references, respectively]

- Abrams, R. A., Meyer, D. E. & Kornblum, S. (1989) Speed and accuracy of saccadic eye movements: Characteristics of impulse variability in the oculomotor system. *Journal of Experimental Psychology: Human Perception and Performance* 15 (3):529–43. [DEK]
- Aizawa, K. (2003) *The systematicity arguments*. Kluwer Academic. [SP]
- Albert, M. C. (1973) A simple test of visual neglect. *Neurology* 23:658–64. [JC]
- Algom, D., Eidels, A., Hawkins, R. X. D., Jefferson, B. & Townsend, J. T. (2015) Features of response times: Identification of cognitive mechanisms through mathematical modeling. In: *The Oxford handbook of computational and mathematical psychology* (Ch. 4), ed. J. Busemeyer, Z. Wang, J. T. Townsend & A. Eidels, pp. 63–98. Oxford University Press. [DRL]
- Allport, A., Meyer, D. E. & Kornblum, S. (1993) Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years. In: *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience*, ed. D. E. Meyer & S. Kornblum, pp. 183–218. MIT Press. [LI]
- Allport, D. A. (1980) Attention and performance. In: *Cognitive psychology*, ed. G. Claxton, pp. 112–53. Routledge & Kegan Paul. [rJH]
- Anderson, J. R. & Lebiere, C. (1998) *The atomic components of thought*. Erlbaum. [DEK, rJH]
- Anderson, M. L. (2010) Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences* 33:245–66. [AP]
- Andrews, L. S., Watson, D. G., Humphreys, G. W. & Braithwaite, J. J. (2011) Flexible feature-based inhibition in visual search mediates magnified impairments of selection: Evidence from carry-over effects under dynamic preview-search conditions. *Journal of Experimental Psychology: Human Perception and Performance* 37:1007–16. [DGW]
- Anstis, S. M. (1974) A chart demonstrating variations in acuity with retinal position. *Vision Research* 14(7):589–92. [DEK]
- Anton-Erxleben, K. & Carrasco, M. (2013) Attentional enhancement of spatial resolution: Linking behavioural and neurophysiological evidence. *Nature Reviews Neuroscience* 14(3):188–200. [AKr]
- Ásgeirsson, Á. G. & Kristjánsson, Á. (2011) Episodic retrieval and feature facilitation in intertrial priming of visual search. *Attention, Perception, and Psychophysics* 73:1350–60. [AKr]
- Aspinall, P. A., Borooah, S., Al Alouch, C., Roe, J., Laude, A., Gupta, R., Gupta, M., Montarzano, A. & Dhillon, B. (2014) Gaze and pupil changes during navigation in age-related macular degeneration. *British Journal of Ophthalmology* 98:1393–97. [DPC]
- Bacon, W. F. & Egeth, H. E. (1994) Overriding stimulus-driven attentional capture. *Perception and Psychophysics* 55:485–96. doi: 10.3758/BF03205306. [aJH]
- Balas, B., Nakano, L. & Rosenholtz, R. (2009) A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision* 9(12):13, 1–18. [AL]
- Ball, K. K., Beard, B. L., Roenker, D. L., Miller, R. L. & Griggs, D. S. (1988) Age and visual search: Expanding the useful field of view. *Journal of the Optical Society of America, A, Optics, Image, and Science* 5:2210–19. doi: 10.1364/JOSAA.5.002210. [aJH]
- Ballesteros, S., Millar, S. & Reales, J. M. (1998) Symmetry in haptic and in visual shape perception. *Perception and Psychophysics* 60:389–404. [AP]
- Bays, P. M. & Husain, M. (2008) Dynamic shifts of limited working memory resources in human vision. *Science* 321(5890):851–54. doi: 10.1126/science.1158023. [SO]
- Beck, J. (1972) Similarity grouping and peripheral discriminability under uncertainty. *American Journal of Psychology* 85:1–19. doi: 10.2307/1420955. [aJH]

- Beck, J. & Ambler, B. (1973) The effects of concentrated and distributed attention on peripheral acuity. *Perception and Psychophysics* 14:225–30. doi: 10.3758/BF03212381. [aJH]
- Beck, J., Sutter, A. & Ivry, R. (1987) Spatial frequency channels and perceptual grouping in texture segregation. *Computer Vision, Graphics, and Image Processing* 37(2):299–325. [RR]
- Becker, S. I. (2010) The role of target–distractor relationships in guiding attention and the eyes in visual search. *Journal of Experimental Psychology: General* 139:247–65. doi: 10.1037/a0018808. [aJH]
- Becker, S. I. & Ansorge, U. (2013) Higher set sizes in pop-out search displays do not eliminate priming or enhance target selection. *Vision Research* 81:18–28. Available at: <http://doi.org/10.1016/j.visres.2013.01.009>. [AKr]
- Becker, S. I., Ansorge, U. & Horstmann, G. (2009) Can inter-trial priming effects account for the similarity effect in visual search? *Vision Research* 49:1738–56. [SVds]
- Belopolsky, A. V. & Theeuwes, J. (2010) No capture outside the attentional window. *Vision Research* 50:2543–50. doi: 10.1016/j.visres.2010.08.023. [aJH]
- Belopolsky, A. V., Zwaan, L., Theeuwes, J. & Kramer, A. F. (2007) The size of an attentional window modulates attentional capture by color singletons. *Psychonomic Bulletin and Review* 14(5):934–38. Available at: <http://doi.org/10.3758/BF03194124>. [AKr]
- Bertera, J. H. (1988) The effect of simulated scotomas on visual search in normal subjects. *Investigative Ophthalmology and Visual Science* 29(3):470–75. [DPC]
- Berti, A. & Rizzolatti, G. (2002) Coding near and far space. In: *The cognitive and neural bases of spatial neglect*, ed. H. Karnath, D. Milner & G. Vallar, pp. 119–29. Oxford University Press. [JC]
- Bichot, N. P., Heard, M. T., DeGemaro, E. M. & Desimone, R. (2015) A source for feature-based attention in the prefrontal cortex. *Neuron* 88(4):832–44. [LI]
- Bichot, N. P., Rossi, A. F. & Desimone, R. (2005) Parallel and serial neural mechanisms for visual search in macaque area V4. *Science* 308:529–34. [KRC, ME]
- Binello, A., Mannan, S. & Ruddock, K. H. (1995) The characteristics of eye movements during visual search with multi-element stimuli. *Spatial Vision* 9:343–62. doi: 10.1163/156856895X00043. [aJH]
- Birmingham, E., Bischof, W. F. & Kingstone, A. (2008) Social attention and real-world scenes: The roles of action, competition and social content. *Quarterly Journal of Experimental Psychology (Hove)* 61(7):986–98. [LH]
- Borji, A. & Itti, L. (2013) State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(1):185–207. [LI]
- Borji, A., Sihite, D. N. & Itti, L. (2013) Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing* 22(1):55–69. [LI]
- Bouma, H. (1970) Interaction effects in parafoveal letter recognition. *Nature* 226:177–78. [aJH, DEK]
- Bouwmeester, L., Heutink, J. & Lucas, C. (2007) The effect of visual training for patients with visual field defects due to brain damage: A systematic review. *Journal of Neurology, Neurosurgery, and Psychiatry* 78(6):555–64. [DPC]
- Brady, T. F., Konkle, T. & Alvarez, G. A. (2009) Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General* 138:487–502. [RW]
- Brady, T. F., Konkle, T. & Alvarez, G. A. (2011) A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision* 11:1–34. [RW]
- Braisy, N. & Gellatly, A. (2012) *Cognitive psychology*, second edition. Oxford University Press. [rJH]
- Brascamp, J. W., Pels, E. & Kristjánsson, Á. (2011) Priming of pop-out on multiple time scales during visual search. *Vision Research* 51:1972–78. [AKr]
- Bravo, M. J. & Nakayama, K. (1992) The role of attention in different visual-search tasks. *Perception and Psychophysics* 51:465–72. doi: 10.3758/BF03211642. [TT, AL]
- Broadbent, D. E. (1958) *Perception and communication*. Pergamon Press. [rJH]
- Buetti, S., Cronin, D. A., Madison, A. M., Wang, Z. & Lleras, A. (2016) Towards a better understanding of parallel visual processing in human vision: Evidence for exhaustive analysis of visual information. *Journal of Experimental Psychology: General* 145(6):672–707. [AL, rJH]
- Bundesen, C., Habekost, T. & Kyllingsbaek, S. (2005) A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review* 112:291–328. doi: 10.1037/0033-295X.112.2.291. [aJH]
- Buschman, T. J. & Miller, E. K. (2009) Serial, covert shifts of attention during visual search are reflected by the frontal eye fields and correlated with population oscillations. *Neuron* 63:386–96. doi: 10.1016/j.neuron.2009.06.020. [AKh, rJH]
- Butt, T., Crossland, M. D., West, P., Orr, S. W. & Rubin, G. S. (2015) Simulation contact lenses for AMD health state utility values in NICE appraisals: A different reality. *British Journal of Ophthalmology* 99(4):540–44. [DPC]
- Cajar, A., Engbert, R. & Laubrock, J. (2016a) Spatial frequency processing in the central and peripheral visual field during scene viewing. *Vision Research* 127:186–97. doi: 10.1016/j.visres.2016.05.008. [JL]
- Cajar, A., Schneeweiß, P., Engbert, R. & Laubrock, J. (2016b) Coupling of attention and saccades when viewing scenes with central and peripheral degradation. *Journal of Vision* 16(2):8, 1–19. [JL]
- Campion, J. (1987) Apperceptive agnosia: The specification and description of constructs. In: *Visual object processing: A cognitive neuropsychological approach*, ed. G. W. Humphreys & M. J. Riddoch, pp. 197–232. Erlbaum. [JC]
- Campion, J. (1989) Interfacing the laboratory with the real world: A cognitive approach to colour assignment in visual displays. In: *Cognitive ergonomics and human-computer interaction*, ed. J. Long & A. Whitefield, pp. 35–65. Cambridge University Press. [JC]
- Campion, J. (2009) *Consciousness: Paradigm for a science of real minds and brains with special reference to vision*. SOTU Press. [JC]
- Campion, J. (2011) Consciousness as a paradigm for cognitive neuroscience. *History and Philosophy of Psychology* 13(2):24–31. [JC]
- Campion, J. (2014) De-inventing the wheel: How information processing destroyed the science of vision. *Paper presented at the History and Philosophy of Psychology Society's Annual Conference*, April 15, Guildford, United Kingdom. [JC]
- Campion, J., Latto, R. & Smith, Y. M. (1983) Is blindsight an effect of scattered light, spared cortex and near threshold vision? *Behavioral and Brain Sciences* 6:423–48. [JC]
- Carrasco, M. (2011) Visual attention: The past 25 years. *Vision Research* 51:1484–525. doi: 10.1016/j.visres.2011.04.012. [aJH, LI]
- Carrasco, M., Evert, D. L., Chang, E. & Katz, S. M. (1995) The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception and Psychophysics* 57:1241–61. doi: 10.3758/BF03208380. [aJH]
- Carrasco, M., Talgar, C. P. & Cameron, E. L. (2001) Characterizing visual performance fields: Effects of transient covert attention, spatial frequency, eccentricity, task and set size. *Spatial Vision* 15(1):61–75. Available at: <http://doi.org/10.1163/15685680152692015>. [AKr]
- Carrasco, M. & Yeshurun, Y. (1998) The contribution of covert attention to the set-size and eccentricity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance* 24(2):673–92. Available at: <http://doi.org/10.1037/0096-1523.24.2.673>. [AKr]
- Castelhano, M. S. & Henderson, J. M. (2008) Stable individual differences across images in human saccadic eye movements. *Canadian Journal of Experimental Psychology* 62(1):1–14. [LH]
- Cavanagh, P. & Alvarez, G. A. (2005) Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences* 9(7):349–54. Available at: <http://doi.org/10.1016/j.tics.2005.05.009>. [AKr]
- Cavanagh, P., Hunt, A. R., Afraz, A. & Rolfs, M. (2010) Visual stability based on remapping of attention pointers. *Trends in Cognitive Science* 14(4):147–53. [SO]
- Cave, K. R. (2012) FIT: Foundation for an integrative theory. In: *From perception to consciousness: Searching for Anne Treisman*, ed. J. M. Wolfe & L. Robertson, pp. 139–45. Oxford University Press. [KRC]
- Cave, K. R. & Zimmerman, J. M. (1997) Flexibility in spatial attention before and after practice. *Psychological Science* 8:399–403. [KRC]
- Cerf, M., Frady, E. P. & Koch, C. (2009) Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision* 9(12):10 11–15. [LH]
- Chan, L. K. H. & Hayward, W. G. (2013) Visual search. *WIREs Cognitive Science* 4:415–29. doi: 10.1002/wcs.1235. [aJH]
- Chang, H. & Rosenholtz, R. (2014) New exploration of classic search tasks. *Journal of Vision* 14(10):933. [RR]
- Charness, N., Reingold, E. M., Pomplun, M. & Stampe, D. M. (2001) The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory and Cognition* 29:1146–52. Available at: <http://dx.doi.org/10.3758/BF03206384>. [TJC]
- Chen, Z. (1998) Switching attention within and between objects: The role of subjective organization. *Canadian Journal of Experimental Psychology* 52:7–16. [KRC]
- Chen, Z. & Cave, K. R. (2013) Perceptual load vs. dilution: The role of attentional focus, stimulus category, and target predictability. *Frontiers in Psychology* 4:327, 1–14. [KRC]
- Chen, Z. & Cave, K. R. (2014) Constraints on dilution from a narrow attentional zoom reveal how spatial and color cues direct selection. *Vision Research* 101:125–37. [KRC]
- Chen, Z. & Cave, K. R. (2016) Zooming in on the cause of the perceptual load effect in the go/no-go paradigm. *Journal of Experimental Psychology: Human Perception and Performance* 42(8):1072–87. Available at: <http://dx.doi.org/10.1037/xhp0000168>. [KRC]
- Chun, M. M. (2000) Contextual cueing of visual attention. *Trends in Cognitive Sciences* 4:170–78. [DGW]
- Chun, M. M. & Jiang, Y. (1998) Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology* 36(1):28–71. [LI]
- Chun, M. M. & Wolfe, J. M. (1996) Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology* 30:39–78. doi: 10.1006/cogp.1996.0002. [aJH]

- Coeckelbergh, T. R. M., Cornelissen, F. W., Brouwer, W. H. & Kooijman, A. C. (2002) The effect of visual field defects on eye movements and practical fitness to drive. *Vision Research* 42(5):669–77. [DPC]
- Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., Linenweber, M. R., Petersen, S. E., Raichle, M. E., Essen, D. C. V. & Shulman, G. L. (1998) A common network of functional areas for attention and eye movements. *Neuron* 21:761–73. [JL]
- Cornelissen, F. W., Bruin, K. J. & Kooijman, A. C. (2005) The influence of artificial scotomas on eye movements during visual search. *Optometry and Vision Science* 82(1):27–35. [DPC]
- Cowan, N. (2001) The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24:87–114. doi: 10.1017/S0140525X01003922. [TT]
- Crabb, D. P., Smith, N. D., Glen, F. C., Burton, R. & Garway-Heath, D. F. (2013) How does glaucoma look? Patient perception of visual field loss. *Ophthalmology* 120(6):1120–26. [DPC]
- Crabb, D. P., Smith, N. D. & Zhu, H. (2014) What's on TV? Detecting age-related neurodegenerative eye disease using eye movement scanpaths. *Frontiers in Aging Neuroscience* 6:312. [DPC]
- Crandall, B., Klein, G. & Hoffman, R. (2006) *Working minds: A practitioner's guide to cognitive task analysis*. MIT Press. [JC]
- Crick, F. (1984) Function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* 81(14):4586–90. [LI]
- Crouzet, S. M., Kirchner, H. & Thorpe, S. J. (2010) Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision* 10(4):16 11–17. [LH]
- de Groot, F., Huettig, F. & Olivers, C. N. (2016) When meaning matters: The temporal dynamics of semantic influences on visual attention. *Journal of Experimental Psychology: Human Perception and Performance* 42:180–96. doi: 10.1037/xhp0000102. [rjH]
- Desimone, R., Albright, T., Gross, C. & Bruce, C. (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience* 4(8):2051–62. [SVdS]
- Desimone, R. & Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annual Review of Neuroscience* 18(1):193–222. [LI, WBUP]
- Deubel, H. & Schneider, W. X. (1996) Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research* 36:1827–37. [KRC]
- Devue, C., Belopolsky, A. V. & Theeuwes, J. (2012) Oculomotor guidance and capture by irrelevant faces. *PLoS ONE* 7(4):e34598. [LH]
- Di Lollo, V. (2012) The feature-binding problem is an ill-posed problem. *Trends in Cognitive Sciences* 16:317–21. doi: 10.1016/j.tics.2012.04.007. [rjH]
- Diaz, G. J., Fajen, B. R. & Phillips, F. (2012) Anticipation from biological motion: The goalkeeper problem. *Journal of Experimental Psychology: Human Perception and Performance* 38:848–64. [JvdK]
- Dicks, M., Button, C. & Davids, K. (2010) Availability of advance visual information constrains association-football goalkeeping performance during penalty kick. *Perception* 39:1111–24. [JvdK]
- Donnelly, N., Cave, K. R., Welland, M. & Menneer, T. (2006) Breast screening, chicken sexing and the search for oil: Challenges for visual cognition. *Geological Society, London, Special Publications* 254:43–55. doi: 10.1144/GSL.SP.2006.254.01.04. [aJH]
- Donovan, T., Manning, D. J., Phillips, P. W., Higham, S. & Crawford, T. (2005) The effect of feedback on performance in a fracture detection task. *SPIE* 5749:79–85. Available at: <http://dx.doi.org/10.1117/12.593294>. [TJC]
- Downing, C. J. (1988) Expectancy and visual-spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance* 14(2):188–202. [WBUP]
- Drew, T., Evans, K. K., Vö, M. L. H., Jacobson, F. L. & Wolfe, J. M. (2013a) Informatics in radiology: What can you see in a single glance and how might this guide visual search in medical images? *Radiographics* 33:263–74. doi: 10.1148/rg.331125023. [aJH]
- Drew, T., Vö, M. L. H., Olwal, A., Jacobson, F., Seltzer, S. E. & Wolfe, J. M. (2013b) Scanners and drillers: Characterizing expert visual search through volumetric images. *Journal of Vision* 13(10):3. doi: 10.1167/13.10.3. [aJH]
- Drew, T., Vö, M. L. H. & Wolfe, J. M. (2013c) The invisible gorilla strikes again: Sustained inattention blindness in expert observers. *Psychological Science* 24:1848–53. doi: 10.1177/0956797613479386. [aJH]
- Driver, J. & Frith, C. (2000) Shifting baselines in attention research. *Nature Reviews Neuroscience* 1(2):147–48. [LI]
- Duncan, J. (1984) Selective attention and the organization of visual information. *Journal of Experimental Psychology: General* 113:501–17. doi: 10.1037/0096-3445.113.4.501. [rjH]
- Duncan, J. (1985) Visual search and visual attention. In: *Attention and performance XI: Attention and neuropsychology*, ed. M. I. Posner & O. S. M. Marin, pp. 85–106. Erlbaum. [aJH]
- Duncan, J. (2006) EPS Mid-Career Award 2004: Brain mechanisms of attention. *Quarterly Journal of Experimental Psychology* 59:2–27. [ME]
- Duncan, J. & Humphreys, G. (1992) Beyond the search surface: Visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception and Performance* 18:578–88. doi: 10.1037/0096-1523.18.2.578. [aJH]
- Duncan, J. & Humphreys, G. W. (1989) Visual search and stimulus similarity. *Psychological Review* 96:433–58. doi: 10.1037/0033-295X.96.3.433. [arJH, HJM, AP, SP]
- Duncan, J., Ward, R. & Shapiro, K. (1994) Direct measurement of attentional dwell time in human vision. *Nature* 369:313–15. doi: 10.1038/369313a0. [arJH]
- Eckstein, M. P. (1998) The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science* 9:111–18. [AKr]
- Eckstein, M. P. (2011) Visual search: A retrospective. *Journal of Vision* 11(5):14. doi: 10.1167/11.5.14. [arJH]
- Eckstein, M. P., Thomas, J. P., Palmer, J. & Shimozaki, S. S. (2000) A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception and Psychophysics* 62:425–51. doi: 10.3758/BF03212096. [arJH]
- Egry, R., Driver, J. & Rafal, R. D. (1994) Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General* 123:161–77. doi: 10.1037/0096-3445.123.2.161. [KRC, SVdS, WBUP, rjH]
- Egner, T. & Hirsch, J. (2005) Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience* 8:1784–90. doi: 10.1038/nn1594. [TT]
- Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A. & Oliva, A. (2009) Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition* 17(6–7):945–78. [LI]
- Ehinger, K. A. & Rosenholtz, R. (2016) A general account of peripheral encoding also predicts scene perception performance. *Journal of Vision* 16(2):13. [RR, rjH]
- Eidels, A., Houpt, J. W., Altieri, N., Pei, L. & Townsend, J. T. (2011) Nice guys finish fast and bad guys finish last: Facilitatory vs. inhibitory interaction in parallel systems. *Journal of Mathematical Psychology* 55:176–90. [DRL]
- Eimer, M. (1996) The N2pc as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology* 99:225–34. doi: 10.1016/0013-4694(96)95711-9. [TT]
- Eimer, M. (2014) The neural basis of attentional control in visual search. *Trends in Cognitive Sciences* 99:225–34. [ME]
- Eimer, M. (2015) EPS Mid-Career Award 2014: The control of attention in visual search: Cognitive and neural mechanisms. *Quarterly Journal of Experimental Psychology* 68:2437–63. doi: 10.1080/17470218.2015.1065283. [aJH, ME]
- Eimer, M. & Grubert, A. (2014) Spatial attention can be allocated rapidly and in parallel to new visual objects. *Current Biology* 24:193–98. doi: 10.1016/j.cub.2013.12.001. [ME, rjH]
- Einhäuser, W., Spain, M. & Perona, P. (2008) Objects predict fixations better than early saliency. *Journal of Vision* 8(14):18. [SVdS]
- Engbert, R. & Kliegl, R. (2003) Microsaccades uncover the orientation of covert attention. *Vision Research* 43(9):1035–45. doi: 10.1016/S0042-6989(03)00084-1. [SO]
- Engbert, R., Mergenthaler, K., Sinn, P. & Pikovsky, A. (2011) An integrated model of fixational eye movements and microsaccades. *Proceedings of the National Academy of Sciences of the United States of America* 108:E765–70. [JL]
- Engbert, R., Trukenbrod, H. A., Barthelmé, S. & Wichmann, F. A. (2015) Spatial statistics and attentional dynamics in scene viewing. *Journal of Vision* 15(1):14, 1–17. [JL]
- Engel, F. L. (1971) Visual conspicuity, directed attention and retinal locus. *Vision Research* 11:563–76. doi: 10.1016/0042-6989(71)90077-0. [rjH]
- Engel, F. L. (1977) Visual conspicuity, visual search and fixation tendencies of the eye. *Vision Research* 17:95–108. doi: 10.1016/0042-6989(77)90207-3. [arJH, DEK]
- Emms, J. T. & Kingstone, A. (1995) Access to global and local properties in visual search for compound stimuli. *Psychological Science* 6:283–91. [JTE]
- Eriksen, C. W. & St. James, J. D. (1986) Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics* 40:225–40. [KRC, TM]
- Ester, E. F., Klee, D. & Awh, E. (2013) Visual crowding cannot be wholly explained by feature pooling. *Journal of Experimental Psychology: Human Perception and Performance* 40(3):1022–33. [AL]
- Ester, E. F., Zilber, E. & Serences, J. T. (2015) Substitution and pooling in visual crowding induced by similar and dissimilar distractors. *Journal of Vision* 15(1):4, 1–12. [AL]
- Evans, K. K., Birdwell, R. L. & Wolfe, J. M. (2013a) 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(5):e64366. doi: 10.1371/journal.pone.0064366. [aJH]
- Evans, K. K., Georgian-Smith, D., Tambouret, R., Birdwell, R. L. & Wolfe, J. M. (2013b) The gist of the abnormal: Above-chance medical decision making in the blink of an eye. *Psychonomic Bulletin and Review* 20(6):1170–75. doi: 10.3758/s13423-013-0459-3. [JMW]

- Eysenck, M. W. & Keane, M. T. (2015) *Cognitive psychology. A student's handbook*, seventh edition. Psychology Press. [rjH]
- Fecteau, J. H., Enns, J. T. & Kingstone, A. (2000) Competition-induced visual field differences in search. *Psychological Science* 11:386–93. [JTE]
- Fific, M., Little, D. R. & Nosofsky, R. (2010) Logical-rule models of classification response times: A synthesis of mental-architecture, random-walk, and decision-bound approaches. *Psychological Review* 117:309–48. [DRL]
- Fific, M., Townsend, J. T. & Eidels, A. (2008) Studying visual search using systems factorial methodology with target-distractor similarity as the factor. *Perception and Psychophysics* 70:583–603. [DRL]
- Findlay, J. M. (1997) Saccade target selection during visual search. *Vision Research* 37:617–31. doi: 10.1016/S0042-6989(96)00218-0. [aJH]
- Findlay, J. M. & Gilchrist, I. D. (1998) Eye guidance and visual search. In: *Eye guidance in reading, driving and scene perception*, ed. G. Underwood, pp. 295–312. Elsevier. [arjH]
- Findlay, J. M. & Gilchrist, I. D. (2001) Visual attention: The active vision perspective. In: *Vision and attention*, ed. M. Jenkins & L. Harris, pp. 85–105. Springer. [arjH]
- Findlay, J. M. & Gilchrist, I. D. (2003) *Active vision: The psychology of looking and seeing*. Oxford University Press. [DEK, rjH]
- Findlay, J. M. & Gilchrist, I. D. (2005) Eye guidance and visual search. In: *Cognitive Processes in Eye Guidance*, ed. G. Underwood, pp. 259–81. Oxford University Press. [arjH]
- Findlay, J. M. & Walker, R. (1999) A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences* 22 (04):661–74. [ZS]
- Fischer, B. (1987) The preparation of visually guided saccades. *Reviews of Physiology Biochemistry and Pharmacology* 106:1–35. [AKh]
- Fiser, J. & Aslin, R. N. (2001) Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science* 12:499–504. [RW]
- Fiser, J. & Aslin, R. N. (2002) Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America* 99(24):15822–26. [RW]
- Fleetwood, M. D. & Byrne, M. D. (2006) Modeling the visual search of displays: A revised ACT-R/PM model of icon search based on eye tracking data. *Human-Computer Interaction* 21(2):153–97. [DEK]
- Fodor, J. A. & Pylyshyn, Z. W. (1988) Connectionism and cognitive architecture: A critical analysis. *Cognition* 28(1–2):3–71. [SP]
- Foley, J. M. & Schwarz, W. (1998) Spatial attention: Effect of position uncertainty and number of distractor patterns on the threshold-versus-contrast function for contrast discrimination. *Journal of the Optical Society of America A* 15(5):1036. Available at: <http://doi.org/10.1364/JOSAA.15.001036>. [AKr]
- Friedman, H. S., Zhou, H. & Von der Heydt, R. (2003) The coding of uniform colour figures in monkey visual cortex. *Journal of Physiology* 548:593–613. doi: 10.1113/jphysiol.2002.033555. [rjH]
- Fuggetta, G., Lanfranchi, S. & Campana, G. (2009) Attention has memory: Priming for the size of the attentional focus. *Spatial Vision* 22(2):147–59. [AKr]
- Geisler, W. S. & Chou, K. L. (1995) Separation of low-level and high-level factors in complex tasks: Visual search. *Psychological Review* 102:356–78. doi: 10.1037/0033-295X.102.2.356. [arjH]
- Gentner, D. (1983) Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7(2):47–59. [SP]
- Geringswald, F., Baumgartner, F. & Pollman, S. (2012) Simulated loss of foveal vision eliminates visual search advantage in repeated displays. *Frontiers in Human Neuroscience* 11(6):134. [DPC]
- Geringswald, F., Herbig, A., Hoffman, M. B. & Pollman, S. (2013) Contextual cueing impairment in patients with age-related macular degeneration. *Journal of Vision* 13(3):28. [DPC]
- Geruschat, D. R., Hassan, S. E., Turano, K. A., Quigley, H. A. & Congdon, N. G. (2006) Gaze behaviour of the visually impaired during street crossing. *Optometry and Vision Science* 83(8):550–58. [DPC]
- Geyer, T., Von Mühlen, A. & Müller, H. J. (2007) What do eye movements reveal about the role of memory in visual search? *Quarterly Journal of Experimental Psychology* 60(7):924–35. [ZS]
- Gibson, J. J. (1962) Observations on active touch. *Psychological Review* 69:477–91. [AP]
- Gibson, J. J. (1979) *The ecological approach to visual perception*. Houghton Mifflin. [JC, JvdK]
- Gilad-Gutnick, S. & Sinha, P. (2012) Recognizing degraded faces: The contribution of configural and featural cues. *Perception* 41(12):1497–511. [AKh]
- Gilchrist, I. D. & Harvey, M. (2000) Refixation frequency and memory mechanisms in visual search. *Current Biology* 10:1209–12. doi: 10.1016/S0960-9822(00)00729-6. [aJH]
- Glen, F. C., Smith, N. D. & Crabb, D. P. (2013) Saccadic eye movements and face recognition performance in patients with central glaucomatous visual field defects. *Vision Research* 82:42–51. [DPC]
- Glen, F. C., Smith, N. D. & Crabb, D. P. (2015) Impact of superior and inferior visual field loss on hazard detection in a computer-based driving test. *British Journal of Ophthalmology* 99(5):613–17. [DPC]
- Godwin, H. J., Menneer, T., Cave, K. R. & Donnelly, N. (2010) Dual-target search for high and low prevalence X-ray threat targets. *Visual Cognition* 18:1439–63. doi: 10.1080/13506285.2010.500605. [aJH]
- Godwin, H. J., Menneer, T., Cave, K. R., Thaibsyah, M. & Donnelly, N. (2015a) The effects of increasing target prevalence on information-processing during visual search. *Psychonomic Bulletin and Review* 22(2):469–75. [TM]
- Godwin, H. J., Menneer, T., Riggs, C. A., Cave, K. R. & Donnelly, N. (2015b) Perceptual failures in the selection and identification of low-prevalence targets in relative prevalence visual search. *Attention, Perception, and Psychophysics* 77:150–59. [TM]
- Godwin, H. J., Reichle, E. D. & Menneer, T. (2014) Coarse-to-fine eye movement behavior during visual search. *Psychonomic Bulletin and Review* 21:1244–49. [TM]
- Godwin, H. J., Reichle, E. D. & Menneer, T. (in press) Modeling lag-2 revisits to understand trade-offs in mixed control of fixation termination during visual search. *Cognitive Science*. [TM]
- Goldsmith, M. & Yeari, M. (2003) Modulation of object-based attention by spatial focus under endogenous and exogenous orienting. *Journal of Experimental Psychology: Human Perception and Performance* 18:26–28. [KRC]
- Goldstein, E. B. (2014) *Sensation and perception*, ninth edition. Wadsworth Cengage. [rjH]
- Goldstein, E. B. (2015) *Cognitive psychology: Connecting mind, research and everyday experience*, fourth edition. Wadsworth Cengage. [rjH]
- Gordon, J. & Abramov, I. (1977) Color vision in the peripheral retina. II. Hue and saturation. *Journal of the Optical Society of America* 67(2):202–207. [DEK]
- Grabowecky, M., List, A., Iordanescu, L. & Suzuki, S. (2011) Haptic influence on visual search. *i-Perception* 2:824. [AP]
- Graham, N., Beck, J. & Sutter, A. (1992) Nonlinear processes in spatial-frequency channel models of perceived texture segregation: Effects of sign and amount of contrast. *Vision Research* 32(4):719–43. [RR]
- Greenberg, A. S., Rosen, M., Cutrone, E. & Behrmann, M. (2015) The effects of visual search efficiency on object-based attention. *Attention, Perception, and Psychophysics* 77(5):1544–57. [WBUP]
- Greene, M. R. & Oliva, A. (2009) The briefest of glances: The time course of natural scene understanding. *Psychological Science* 20:464–72. doi: 10.1111/j.1467-9280.2009.02316.x. [aJH]
- Greenwood, J. A., Bex, P. J. & Dakin, S. C. (2009) Positional averaging explains crowding with letter-like stimuli. *Proceedings of the National Academy of Sciences of the United States of America* 106(31):13130–35. [RR]
- Greenwood, J. A., Bex, P. J. & Dakin, S. C. (2012) Crowding follows the binding of relative position and orientation. *Journal of Vision* 12(3):1–20. [RR]
- Grubert, A. & Eimer, M. (2015) Rapid parallel attentional target selection in single-color and multiple-color visual search. *Journal of Experimental Psychology: Human Perception and Performance* 41:86–101. doi: 10.1037/xhp0000019. [rjH]
- Haber, R. N. (1983) The impending demise of the icon: The role of iconic processes in information processing theories of perception. *Behavioral and Brain Sciences* 6:1–11. [aJH]
- Hafed, Z. M. (2011) Mechanisms for generating and compensating for the smallest possible saccades. *European Journal of Neuroscience* 33(11):2101–13. doi: 10.1111/j.1460-9568.2011.07694.x. [SO]
- Hafed, Z. M. & Clark, J. J. (2002) Microsaccades as an overt measure of covert attention shifts. *Vision Research* 42(22):2533–45. [SO]
- Hafed, Z. M., Goffart, L. & Krauzlis, R. J. (2009) A neural mechanism for micro-saccade generation in the primate superior colliculus. *Science* 323(5916):940–43. [SO]
- Hall, M., Mattingley, J. & Dux, P. (2015) Distinct contributions of attention and working memory to visual statistical learning and ensemble processing. *Journal of Experimental Psychology: Human Perception and Performance* 41:1112–23. [RW]
- Halverson, T. & Hornof, A. J. (2011) A computational model of “active vision” for visual search in human-computer interaction. *Human-Computer Interaction* 26 (4):285–314. [DEK]
- Hari, R., Henriksson, L., Malinen, S. & Parkkonen, L. (2015) Centrality of social interaction in human brain function. *Neuron* 88(1):181–93. [LH]
- Harms, L. & Bundesen, C. (1983) Color segregation and selective attention in a nonsearch task. *Perception and Psychophysics* 33:11–19. [KRC]
- Harris, C. M. (1995) Does saccadic undershoot minimize saccadic flight-time? A Monte-Carlo study. *Vision Research* 35:691–701. [DEK]
- Harvey, H. & Walker, R. (2014) Reading with peripheral vision: A comparison of reading dynamic scrolling and static text with a simulated central scotoma. *Vision Research* 98:54–60. [DPC]
- He, K., Zhang, X., Ren, S. & Sun, J. (2015) Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. *arXiv* 1502.01852. [SVdS]
- He, S., Cavanagh, P. & Intriligator, J. (1996) Attentional resolution and the locus of visual awareness. *Nature* 383:334–37. doi: 10.1038/383334a0. [arjH]
- He, Z. J. & Nakayama, K. (1992) Surfaces versus features in visual search. *Nature* 359:231–33. [AP]

- Heller, M. A. (1984) Active and passive touch: The influence of exploration time on form recognition. *Journal of General Psychology* 110:243–49. [AP]
- Henderson, J. M. & Castellano, M. S. (2005) Eye movements and visual memory for scenes. In: *Cognitive processes in eye guidance*, ed. G. Underwood, pp. 213–35. Oxford University Press. [DEK]
- Henderson, J. M. & Ferreira, F. (1990) Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16:417–29. [JL]
- Henderson, J. M. & Hollingworth, A. (1999) The role of fixation position in detecting scene changes across saccades. *Psychological Science* 10:438–43. doi: 10.1111/1467-9280.00183. [aJH]
- Hickey, C., Chelazzi, L. & Theeuwes, J. (2010) Reward has a larger impact on visual search in people with reward-seeking personalities. *Journal of Vision* 10:255–55. [AP]
- Hochberg, J. (1968) In the mind's eye. In: *Contemporary theory and research in visual perception*, ed. R. N. Haber, pp. 309–31. Holt Rinehart and Winston. [JTE]
- Hochberg, J. (1982) How big is a stimulus? In: *Contemporary theory and research in visual perception*, ed. J. Beck, pp. 191–217. Erlbaum. [JTE, rJH]
- Hochstein, S. & Ahissar, M. (2002) View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron* 36:791–804. doi: 10.1016/S0896-6273(02)01091-7. [rJH]
- Hoffman, J. E. & Nelson, B. (1981) Spatial selectivity in visual search. *Perception and Psychophysics* 30:283–90. [KRC]
- Hooge, I. T. C. & Erkelens, C. J. (1996) Control of fixation duration in a simple search task. *Perception and Psychophysics* 58:969–76. doi: 10.3758/BF03206825. [aJH]
- Hooge, I. T. C. & Erkelens, C. J. (1998) Adjustment of fixation duration in visual search. *Vision Research* 38(9):1295–302. doi: 10.1016/S0042-6989(97)00287-3. [ZS]
- Hornof, A. J. (2004) Cognitive strategies for the visual search of hierarchical computer displays. *Human-Computer Interaction* 19(3):183–223. [DEK]
- Horowitz, T. S. & Wolfe, J. M. (1998) Visual search has no memory. *Nature* 394:575–77. doi: 10.1038/29068. [aJH]
- Hout, J. W. & Townsend, J. T. (2012) Statistical measures for workload capacity analysis. *Journal of Mathematical Psychology* 56:341–55. [DRL]
- Hout, M. C., Walenchok, S. C., Goldinger, S. D. & Wolfe, J. M. (2015) Failures of perception in the low-prevalence effect: Evidence from active and passive visual search. *Journal of Experimental Psychology: Human Perception and Performance*. 41:977–94. [TM]
- Hsiao, J. H. & Cottrell, G. (2008) Two fixations suffice in face recognition. *Psychological Science* 19(10):998–1006. [LH]
- Hubel, D. H. & Wiesel, T. N. (1959) Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology* 148(3):574–91. [SVdS]
- Hubel, D. H. & Wiesel, T. N. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology (London)* 160:106–54. [rJH]
- Hubel, D. H. & Wiesel, T. N. (1977) Functional architecture of macaque visual cortex. *Proceedings of the Royal Society Series B: Biological Sciences* 198:1–59. doi: 10.1098/rspb.1977.0085. [rJH]
- Hulleman, J. (2009) No need for inhibitory tagging of locations in visual search. *Psychonomic Bulletin and Review* 16:116–20. doi: 10.3758/PBR.16.1.116. [arJH]
- Hulleman, J. (2010) Inhibitory tagging in visual search: Only in difficult search are items tagged individually. *Vision Research* 50:2069–79. doi: 10.1016/j.visres.2010.07.017. [arJH]
- Hulleman, J. & Olivers, C. N. L. (2014) Search through complex motion displays does not break down under spatial memory load. *Psychonomic Bulletin and Review* 21:652–58. doi: 10.3758/s13423-013-0537-6. [arJH]
- Hulleman, J., Te Winkel, W. & Boselie, F. (2000) Concavities as basic features in visual search: Evidence from search asymmetries. *Perception and Psychophysics* 62:162–74. doi: 10.3758/BF03212069. [aJH]
- Humphreys, G. W. & Müller, H. J. (1993) SEarch via Recursive Rejection (SERR): A connectionist model of visual search. *Cognitive Psychology* 25:43–110. doi: 10.1006/cogp.1993.1002. [arJH, HJM]
- Husemoller, D. (1994) *Fibre bundles*, third edition. Springer-Verlag. [SP]
- Ikeda, M. & Takeuchi T. (1975) Influence of foveal load on the functional visual field. *Perception and Psychophysics* 18:255–60. doi: 10.3758/BF03199371. [rJH]
- Intriligator, J. & Cavanagh, P. (2001) The spatial resolution of visual attention. *Cognitive Psychology* 43:171–216. doi: 10.1006/cogp.2001.0755. [aJH, RR]
- Irwin, D. E. (1991) Information integration across saccadic eye movements. *Cognitive Psychology* 23(3):420–56. [SO]
- Itti, L. & Borji, A. (2013) Computational models: Bottom-up and top-down aspects. In: *The Oxford handbook of attention*, ed. K. Nobre & S. Kastner, pp. 1122–58. Oxford University Press. [LI]
- Itti, L. & Koch, C. (2001) Computational modelling of visual attention. *Nature Reviews Neuroscience* 2(3):194–203. [LI, HJM]
- Itti, L. & Koch, C. (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research* 40:1489–506. doi: 10.1016/S0042-6989(99)00163-7. [aJH]
- Jacko, J. A., Barreto, A. B., Chu, J. Y. M., Scott, I. U., Rosa Jr, R. H. & Pappas Jr, C. C. (2000) Macular degeneration and visual search: What we can learn from eye movement analysis. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 44(29):116–19. [DPC]
- Jacko, J. A., Scott, I. U., Barreto, A. B., Bartsch, H. S., Chu, J. Y. M. & Fain, W. B. (2001) Iconic visual search strategies: A comparison of computer users with AMD versus computer users with normal vision. In: *Proceedings of the Ninth International Conference on Human-Computer Interaction* 2:423. [DPC]
- Jacobs, A. M. (1986) Eye-movement control in visual search: How direct is visual span control. *Perception and Psychophysics* 39:47–58. doi: 10.3758/BF03207583. [aJH]
- Jefferies, L. N., Enns, J. T. & Di Lollo, V. (2014) The flexible focus: Whether spatial attention is unitary or divided depends on observer goals. *Journal of Experimental Psychology: Human Perception and Performance* 40(2):465–70. Available at: <http://doi.org/10.1037/a0034734>. [AKr]
- Johnson, M., Rosebrugh, R. & Wood, R. (2012) Lenses, fibrations and universal translations. *Mathematical Structures in Computer Science* 22:25–42. [SP]
- Jones, M. & Dzhaferov, E. N. (2014) Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review* 121(1):1–32. [RM]
- Jonides, J. & Gleitman, H. (1972) A conceptual category effect of visual search: O as letter or as digit. *Perception and Psychophysics* 12:457–60. doi: 10.3758/BF03210934. [aJH]
- Jonikaitis, D., Szinte, M., Rolf, M. & Cavanagh, P. (2013) Allocation of attention across saccades. *Journal of Neurophysiology* 109(5):1425–34. [SO]
- Kahneman, D. & Treisman, A. (1984) Changing views of attention and automaticity. In: *Varieties of attention*, ed. R. Parasuraman & D. R. Davies, pp. 29–61. Academic Press. [AKh]
- Kaptein, N. A., Theeuwes, J. & Van der Heijden, A. H. C. (1995) Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception and Performance* 21:1053–69. doi: 10.1037/0096-1523.21.5.1053. [aJH]
- Karnath, H., Milner, D. & Vallar, G. (2002) *The cognitive and neural bases of spatial neglect*. Oxford University Press. [JC]
- Kerzel, D., Born, S. & Schönhammer, J. (2012) Perceptual grouping allows for attention to cover noncontiguous locations and suppress capture from nearby locations. *Journal of Experimental Psychology: Human Perception and Performance* 38(6):1362–70. Available at: <http://doi.org/10.1037/a0027780>. [AKr]
- Keshvari, S. & Rosenholtz, R. (2016) Pooling of continuous features provides a unifying account of crowding. *Journal of Vision* 16(3):39, 1–15. doi: 10.1167/16.3.39. [RR, rJH]
- Kieras, D. (2011) The persistent visual store as the locus of fixation memory in visual search tasks. *Cognitive Systems Research* 12:102–12. [DEK]
- Kieras, D. E. & Hornof, A. J. (2014) Towards accurate and practical predictive models of active-vision-based visual search. In: *CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3875–84. ACM. [DEK]
- Kieras, D. E. & Meyer, D. E. (1997) An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction* 12(4):391–438. [DEK]
- Kieras, D. E. & Meyer, D. E. (2000) The role of cognitive task analysis in the application of predictive models of human performance. In: *Cognitive task analysis*, ed. J. M. C. Schraagen, S. E. Chipman & V. L. Shalin, pp. 237–60. Erlbaum. [DEK]
- Kim, M. S. & Cave, K. R. (1995) Spatial attention in visual search for features and feature conjunctions. *Psychological Science* 6:376–80. [KRC]
- Kirkham, N. Z., Slemmer, J. A. & Johnson, S. P. (2002) Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition* 83:B35–42. [RW]
- Klein, R. M. (1988) Inhibitory tagging system facilitates visual search. *Nature* 334:430–31. doi: 10.1038/334430a0. [aJH]
- Klein, R. M. & Farrell, M. (1989) Search performance without eye-movements. *Perception and Psychophysics* 46:476–82. doi: 10.3758/BF03210863. [aJH]
- Ko, H.-K., Poletti, M. & Rucci, M. (2010) Microsaccades precisely relocate gaze in a high visual acuity task. *Nature Neuroscience* 13(12):1549–53. doi: 10.1038/nn.2663. [SO]
- Koffka, K. (1935) *Principles of Gestalt psychology*. Harcourt, Brace. [JC]
- Kraiss, K. F. & Kräuper, A. (1982) Using visual lobe area measurements to predict visual search performance. *Human Factors* 24:673–82. [rJH]
- Kristjánsson, Á. (2015) Reconsidering visual search. *i-Perception* 6(6):1–14. [AKr]
- Kristjánsson, Á. & Campana, G. (2010) Where perception meets memory: A review of priming in visual search. *Attention, Perception, and Psychophysics* 72:5–18. [AKr]

- Kristjánsson, Á. & Driver, J. (2008) Priming in visual search: Separating the effects of target repetition, distractor repetition and role-reversal. *Vision Research* 48 (10):1217–32. Available at: <http://doi.org/10.1016/j.visres.2008.02.007>. [AKr, SvDs]
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012) ImageNet classification with deep convolutional neural networks. In: *Conference of Advances in neural information processing systems*, ed. F. Pereira, C. J. C. Burges, L. Bottou & K. Q. Weinberger, pp. 1097–105. Neural Information Processing Systems Foundation. Available at: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks>. [LH, SvDs]
- Kruijne, W., Brascamp, J. W., Kristjánsson, Á. & Meeter, M. (2015) Can a single short-term mechanism account for priming of pop-out? *Vision Research* 115:17–22. [AKr]
- Kruijne, W. & Meeter, M. (2015) The long and the short of priming in visual search. *Attention, Perception, and Psychophysics* 77(5):1558–73. [AKr]
- Kujala, M. V., Kujala, J., Carlson, S. & Hari, R. (2012) Dog experts' brains distinguish socially relevant body postures similarly in dogs and humans. *PLoS ONE* 7(6): e39145. [LH]
- Kümmerer, M., Theis, L. & Bethge, M. (2014) Deep Gaze I: Boosting saliency prediction with feature maps trained on ImageNet. arXiv preprint arXiv:1411.1045. Available at: <http://arxiv.org/abs/1411.1045>. [LH]
- Kundel, H. L., Nodine, C. F., Conant, E. F. & Weinstein, S. P. (2007) Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology* 242(2):396–402. [LH]
- Kundel, H. L., Nodine, C. F. & Toto, L. C. (1984) Eye movements and the detection of lung tumors in chest images. In: *Theoretical and applied aspects of eye movement research*, ed. A. G. Gale & F. Johnson, pp. 297–304. Elsevier. Available at: [http://dx.doi.org/10.1016/S0166-4115\(08\)61846-4](http://dx.doi.org/10.1016/S0166-4115(08)61846-4). [TJC]
- Kusunoki, M. & Goldberg, M. E. (2003) The time course of perisaccadic receptive field shifts in the lateral intraparietal area of the monkey. *Journal of Neurophysiology* 89:1519–27. doi: 10.1152/jn.00519.2002. [aJH]
- Kusunoki, M., Gottlieb, J. & Goldberg, M. E. (2000) The lateral intraparietal area as a saliency map: The representation of abrupt onset, stimulus motion, and task relevance. *Vision Research* 40(10):1459–68. [LI]
- Kuyk, T. K., Liu, L., Elliott, J. & Fuhr, P. S. W. (2010) Visual search training and obstacle avoidance in adults with visual impairments. *Journal of Visual Impairment and Blindness* 104(4):215–27. [DPC]
- Kuyk, T. K., Liu, L. & Fuhr, P. S. W. (2005) Feature search in persons with severe visual impairment. *Vision Research* 45(25–26):3224–34. [DPC]
- LaBerge, D. & Brown V. (1986) Variations in size of the visual field in which targets are presented: An attentional range effect. *Perception and Psychophysics* 40:188–200. doi: 10.3758/BF03203016. [rJH]
- Lacey, S., Tal, N., Amedi, A. & Sathian, K. (2009) A putative model of multisensory object representation. *Brain Topography* 21:269–74. [AP]
- Lacretie, A. & Frigaszy, D. M. (1997) Manual exploratory procedures and asymmetries for a haptic search task: A comparison between capuchins (*Cebus apella*) and humans. *Laterality: Asymmetries of Body, Brain and Cognition* 2:247–66. [AP]
- Lamme, V. A. F. & Roelfsema, P. R. (2000) The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences* 23:571–79. doi: 10.1016/S0166-2236(00)01657-X. [rJH]
- Lamoureux, E. L., Pallant, J. F., Pesudovs, K., Tennant, A., Rees, G., O'Connor, P. M. & Keefe, J. E. (2008) Assessing participation in daily living and the effectiveness of rehabilitation in age related macular degeneration patients using the impact of vision impairment scale. *Ophthalmic Epidemiology* 15 (2):105–13. [DPC]
- Lamy, D. & Kristjánsson, A. (2013) Is goal-directed attentional guidance just inter-trial priming? A Review. *Journal of Vision* 13(3):14, 1–19. [AL]
- Laubrock, J., Cajar, A. & Engbert, R. (2013) Control of fixation duration during scene viewing by interaction of foveal and peripheral processing. *Journal of Vision* 13(12):11, 1–20. [JL]
- Laubrock, J., Engbert, R. & Kliegl, R. (2005) Microsaccade dynamics during covert attention. *Vision Research* 45:721–30. [JL]
- Laubrock, J., Engbert, R. & Kliegl, R. (2008) Fixational eye movements predict the perceived direction of ambiguous apparent motion. *Journal of Vision* 8(14):13, 1–17. [JL]
- Lavie, N. (1995) Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance* 21 (3):451–68. doi: 10.1037/0096-1523.21.3.451. [AKh, rJH]
- Lavie, N. (2005) Distracted and confused? Selective attention under load. *Trends in Cognitive Sciences* 9:75–82. [KRC]
- Lavie, N., Hirst, A., De Fockert, J. W. & Viding, E. (2004) Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General* 133(3):339–51. [AKr]
- Lavie, N. & Tsai, Y. (1994) Perceptual load as a major determinant of the locus of selection in visual attention. *Perception and Psychophysics* 56:183–97. doi: 10.3758/BF03213897. [AKh, rJH]
- Le, Q. V., Ranzato, M., Monga, R., Devin, M., Chen, K., Corrado, G. S., Dean, J. & Ng, A. Y. (2012) Building high-level features using large scale unsupervised learning. Paper presented at the ICML. [SvDs]
- LeCun, Y., Kavukcuoglu, K. & Farabet, C. (2010) Convolutional networks and applications in vision. Paper presented at the ISCAS. [SvDs]
- Lederman, S. J. & Klatzky, R. L. (1987) Hand movements: A window into haptic object recognition. *Cognitive Psychology* 19:342–68. [AP]
- Lee, H., Grosse, R., Ranganath, R. & Ng, A. Y. (2009) Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Paper presented at the ACM. [SvDs]
- Levi, D. M. (2008) Crowding: An essential bottleneck for object recognition: A mini-review. *Vision Research* 48:635–54. doi: 10.1016/j.visres.2007.12.009. [aJH]
- Levi, D. M. & Carney, T. (2009) Crowding in peripheral vision: Why bigger is better. *Current Biology* 19(23):1988–93. [RR]
- Li, F. F., VanRullen, R., Koch, C. & Perona, P. (2002) Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America* 99(14):9596–601. [JMW]
- Li, H., Bao, Y., Pöppel, E. & Su, Y. H. (2014) A unique visual rhythm does not pop out. *Cognitive Processes* 15:93–97. doi: 10.1007/s10339-013-0581-1. [aJH]
- Li, W., Piëch, V. & Gilbert, C. D. (2004) Perceptual learning and top-down influences in primary visual cortex. *Nature Neuroscience* 7(6):651–57. [AL]
- Li, W., Piëch, V. & Gilbert, C. D. (2006) Contour saliency in primary visual cortex. *Neuron* 50(6):951–62. [AL]
- Liesefeld, H. R., Moran, R., Usher, M., Müller, H. J. & Zehetleitner, M. (2016) Search efficiency as a function of target saliency: The transition from inefficient to efficient search and beyond. *Journal of Experimental Psychology: Human Perception and Performance* 42(6):821–36. doi: 10.1037/xhp0000156. [RM, HJM, rJH]
- Litchfield, D. & Donovan, T. (2016) Worth a quick look? Initial scene previews can guide eye movements as a function of domain-specific expertise but can also have unforeseen costs. *Journal of Experimental Psychology: Human Perception and Performance* 42(7):982–94. doi: 10.1037/xhp0000202. [TJC]
- Little, D. R., Eidels, A., Fific, M. & Wang, T. (2015) Understanding the influence of distractors on workload capacity. *Journal of Mathematical Psychology* 68:25–36. [DRL]
- Liu, L., Kuyk, T. K. & Fuhr, P. S. W. (2007) Visual search training in subjects with severe to profound low vision. *Vision Research* 47(20):2627–2636. [DPC]
- Liversedge, S. P. & Findlay, J. M. (2000) Eye movements reflect cognitive processes. *Trends in Cognitive Sciences* 4:6–14. [TM]
- Lleras, A., Rensink, R. A. & Enns, J. T. (2005) Rapid resumption of interrupted visual search new insights on the interaction between vision and memory. *Psychological Science* 16(9):684–88. [JTE]
- Lleras, A., Rensink, R. A. & Enns, J. T. (2007) Consequences of display changes during interrupted visual search: Rapid resumption is target specific. *Perception and Psychophysics* 69(6):980–93. [JTE]
- Loschky, L. C. & McConkie, G. W. (2002) Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display. *Journal of Experimental Psychology: Applied* 8:99–117. [JL]
- Loschky, L. C., McConkie, G. W., Yang, J. & Miller, M. E. (2005) The limits of visual resolution in natural scene viewing. *Visual Cognition* 12:1057–92. [JL]
- Loughman, J., Davison, P. & Flitcroft, I. (2007) Open angle glaucoma effects on preattentive visual search efficiency for flicker, motion displacement and orientation pop-out tasks. *British Journal of Ophthalmology* 91(11):1493–98. [DPC]
- Luck, S. J. & Hillyard, S. A. (1994) Electrophysiological correlates of feature analysis during visual search. *Psychophysiology* 31:291–308. doi: 10.1111/j.1469-8986.1994.tb02218.x. [TT]
- Lupyan, G. (2008) The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition* 108:566–77. doi: 10.1016/j.cognition.2008.03.009. [rJH]
- Mac Lane, S. (1998) *Categories for the working mathematician*, second edition. Springer. [SP]
- MacKeben, M. & Fletcher, D. C. (2011) Target search and identification performance in low vision patients. *Investigative Ophthalmology and Visual Science* 52:7603–609. [DPC]
- Mackworth, N. H. (1948) The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology* 1:6–21. doi: 10.1080/17470214808416738. [aJH]
- Malcolm, G. L. & Henderson, J. M. (2009) The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision* 9(11):8, 1–13. [JL]
- Malcolm, G. L. & Henderson, J. M. (2010) Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision* 10(2):4, 1–11. [JL]
- Maljkovic, V. & Nakayama, K. (1994) Priming of pop-out: I. Role of features. *Memory and Cognition* 22(6):657–72. [AKr]
- Mangun, G. R. & Hillyard, S. A. (1995) Mechanisms and models of selective attention. In: *Electrophysiology of mind: Event-related brain potentials and cognition*, ed. M. D. Rugg & M. G. H. Coles, pp. 40–85. Oxford University Press. [KRC]
- Manning, D., Barker-Mill, S. C., Donovan, T. & Crawford, T. (2005) Time-dependent observer errors in pulmonary nodule Detection. *The British Journal of Radiology* 78:1–5. Available at: <http://dx.doi.org/10.1259/bjr/13453920>. [TJC]

- Manning, D., Ethell, S., Donovan, T. & Crawford, T. (2006) How do radiologists do it? The influence of experience and training on searching for chest nodules. *Radiography* 12:134–42. Available at: <http://dx.doi.org/10.1016/j.radi.2005.02.003>. [TJC]
- Marr, D. (1982) *Vision: A computational investigation into the human representation and processing of visual information*. W.H. Freeman. [aJH, JC, SVdS]
- Marti, S., Bayet, L. & Dehaene, S. (2015) Subjective report of eye fixations during serial search. *Consciousness and Cognition* 33:1–15. doi: 10.1016/j.concog.2014.11.007. [AKH]
- Martini, P. (2010) System identification in priming of pop-out. *Vision Research* 50:2110–15. [AKr]
- Mazza, V. & Caramazza, A. (2011) Temporal brain dynamics of multiple object processing: The flexibility of individuation. *PLoS ONE* 6(2):e17453. doi: 10.1371/journal.pone.0017453. [TT]
- Mazza, V., Pagano, S. & Caramazza, A. (2013) Multiple object individuation and exact numeration. *Journal of Cognitive Neuroscience* 25:697–705. doi: 10.1162/jocn_a_00349. [TT]
- Mazza, V., Turatto, M., Umiltà, C. & Eimer, M. (2007) Attentional selection and identification of visual objects are reflected by distinct electrophysiological responses. *Experimental Brain Research* 181:531–36. doi: 10.1007/s00221-007-1002-4. [TT]
- McCarley, J. S., Wang, R. X. F., Kramer, A. F., Irwin, D. E. & Peterson, M. S. (2003) How much memory does oculomotor search have? *Psychological Science* 14:422–26. doi: 10.1111/1467-9280.01457. [aJH]
- McConkie, G. W. & Rayner, K. (1975) The span of the effective stimulus during a fixation in reading. *Perception and Psychophysics* 17:578–86. doi: 10.3758/BF03203972. [JL, TM, rJH]
- McLeod, P., Driver, J. & Crisp, J. (1988) Visual search for a conjunction of motion and form is parallel. *Nature* 332:154–55. doi: 10.1038/332154a0. [aJH]
- Meeter, M. & Van der Stigchel, S. (2013) Visual priming through a boost of the target signal: Evidence from saccadic landing positions. *Attention, Perception, and Psychophysics* 75:1336–41. [SVdS]
- Meeter, M., Van der Stigchel, S. & Theeuwes, J. (2010) A competitive integration model of exogenous and endogenous eye movements. *Biological Cybernetics* 102:271–91. [SVdS]
- Menner, T., Barrett, D. J., Phillips, L., Donnelly, N. & Cave, K. R. (2007) Costs in searching for two targets: Dividing search across target types could improve airport security screening. *Applied Cognitive Psychology* 21:915–32. doi: 10.1002/acp.1305. [aJH]
- Meyer, D. E. & Kieras, D. E. (1997) A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review* 104(1):3–65. doi: 10.1037/0033-295X.104.1.3. [DEK, rJH]
- Meyer, D. E. & Kieras, D. E. (1999) Precise to a practical unified theory of cognition and action: Some lessons from computational modeling of human multiple-task performance. In: *Attention and performance XVII. Cognitive regulation of performance: Integration of theory and application*, ed. D. Gopher & A. Koriat, pp. 17–88. MIT Press. [DEK]
- Mirpour, K. & Bisley, J. W. (2012) Anticipatory remapping of attentional priority across the entire visual field. *Journal of Neuroscience* 32(46):16449–57. [SO]
- Moffitt, K. (1980) Evaluation of the fixation duration in visual search. *Perception and Psychophysics* 27(4):370–72. doi: 10.3758/BF03206127. [ZS]
- Mole, C. & Zhao, J. (2016) Vision and abstraction: An empirical refutation of Nico Orlandi's non-cognitivism. *Philosophical Psychology* 29:365–73. [RW]
- Moore, C. M. & Wolfe, J. M. (2001) Getting beyond the serial/parallel debate in visual search: A hybrid approach. In: *The limits of attention: Temporal constraints on human information processing*, ed. K. Shapiro, pp. 178–98. Oxford University Press. [aJH]
- Moran, R., Zehetleitner, M., Liesefeld, H. R., Müller, H. J., & Usher, M. (2016) Serial vs. parallel models of attention in visual search: accounting for benchmark RT-distributions. *Psychonomic Bulletin and Review* 23:1300–15. doi: 10.3758/s13423-015-0978-1. [RM, HJM]
- Moran, R., Zehetleitner, M., Müller, H. J. & Usher, M. (2013) Competitive guided search: Meeting the challenge of benchmark RT-distributions. *Journal of Vision* 13(8):24. doi: 10.1167/13.8.24. [aJH, RM, HJM]
- Motter, B. C. & Belky, E. J. (1998a) The guidance of eye movements during active visual search. *Vision Research* 38:1805–15. doi: 10.1016/S0042-6989(97)00349-0. [aJH]
- Motter, B. C. & Belky, E. J. (1998b) The zone of focal attention during active visual search. *Vision Research* 38:1007–22. doi: 10.1016/S0042-6989(97)00252-6. [aJH]
- Müller, H. J., Heller, D. & Ziegler, J. (1995) Visual search for singleton feature targets within and across feature dimensions. *Perception and Psychophysics* 57:1–17. doi: 10.3758/BF03211845. [HJM]
- Müller, H. J., Humphreys, G. W. & Donnelly, N. (1994) SEarch via Recursive Rejection (SERR): Visual search for single and dual form-conjunction targets. *Journal of Experimental Psychology: Human Perception and Performance* 20:235–58. doi: 10.1037/0096-1523.20.2.235. [HJM]
- Müller, H. J., Krummenacher, J. & Heller, D. (2004) Dimension-specific inter-trial facilitation in visual search for pop-out targets: Evidence for a top-down modifiable visual short-term memory effect. *Visual Cognition* 11:577–602. doi: 10.1080/13506280344000419. [HJM]
- Müller, H. J., von Mühlen, A. & Geyer, T. (2007) Top-down inhibition of distractors in parallel visual search. *Perception and Psychophysics* 69:1373–88. doi: 10.3758/BF03192953. [HJM]
- Murphy, K. J. St. & Foley-Fisher, J. A. (1988) Visual search with non-foveal vision. *Ophthalmic and Physiological Optics* 8(3):345–48. [DPC]
- Murphy, K. J. St. & Foley-Fisher, J. A. (1989) Effect of a scotoma on eye movements during visual search. *Ophthalmic and Physiological Optics* 9(3):317–21. [DPC]
- Najemnik, J. & Geisler, W. S. (2008) Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision* 8(3):4. doi: 10.1167/8.3.4. [aJH]
- Nakayama, K. (1990) The iconic bottleneck and the tenuous link between early visual processing and perception. In: *Vision: Coding and efficiency*, ed. C. Blakemore, pp. 411–22. Cambridge University Press. [JMW]
- Nakayama, K. & Martini, P. (2011) Situating visual search. *Vision Research* 51:1526–37. doi: 10.1016/j.visres.2010.09.003. [aJH, AKr]
- Nakayama, K. & Silverman, G. H. (1986) Serial and parallel processing of visual feature conjunctions. *Nature* 320:264–65. doi: 10.1038/320264a0. [aJH]
- Nako, R., Smith, T. J. & Eimer, M. (2015) Activation of new attentional templates for real-world objects in visual search. *Journal of Cognitive Neuroscience* 27:902–12. [RW]
- Nako, R., Wu, R. & Eimer, M. (2014a) Rapid guidance of visual search by object categories. *Journal of Experimental Psychology: Human Perception and Performance* 40(1):50–60. [RW]
- Nako, R., Wu, R., Smith, T. J. & Eimer, M. (2014b) Item and category-based attentional control during search for real-world objects: Can you find the pants among the pans? *Journal of Experimental Psychology: Human Perception and Performance* 40(4):1283–88. [RW]
- Neider, M. B. & Zelinsky, G. J. (2006) Scene context guides eye movements during visual search. *Vision Research* 46(5):614–21. [LH]
- Neider, M. B. & Zelinsky, G. J. (2008) Exploring set size effects in scenes: Identifying the objects of search. *Visual Cognition* 16(1):1–10. [aJH, AL]
- Neisser, U. (1963) Decision-time without reaction time: Experiments in visual scanning. *American Journal of Psychology* 76:376–85. [WP]
- Neisser, U. (1964) Visual search. *Scientific American* 210:94–102. doi: 10.1038/scientificamerican0664-94. [aJH]
- Neisser, U. (1966) *Cognitive psychology*. Appleton-Century-Crofts. [JC, WP]
- Neisser, U. (1967) *Cognitive psychology*. Prentice-Hall. [WBUP]
- Neri, P. & Levi, D. M. (2006) Spatial resolution for feature binding is impaired in peripheral and amblyopic vision. *Journal of Neurophysiology* 96(1):142–53. doi: 10.1152/jn.01261.2005. [aJH]
- Newell, F. N., Woods, A. T., Mernagh, M. & Bühlhoff, H. H. (2005) Visual, haptic, and crossmodal recognition of scenes. *Experimental Brain Research* 161:233–42. [AP]
- Nodine, C. F. & Kundel, H. L. (1987) Using eye movements to study visual search and to improve tumor detection. *Radiographics* 7(6):1241–50. [DPC]
- Norman, D. A. & Shallice, T. (1980) *Attention to action: Willed and automatic control of behavior* (Technical Report No. 99). Center for Human Information Processing. [rJH]
- Norman, D. A. & Shallice, T. (1986) Attention to action: Willed and automatic control of behavior. In: *Consciousness and self-regulation*, ed. R. J. Davidson, G. E. Schwartz & D. Shapiro, pp. 1–18. Plenum Press. [rJH]
- Norman, L. J., Heywood, C. A. & Kentridge, R. W. (2013) Object-based attention without awareness. *Psychological Science* 24(6):837–43. doi: 10.1177/0956797612461449. [WBUP]
- Nuthmann, A. (2014) How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance* 40:342–60. [JL]
- Nuthmann, A., Smith, T. J., Engbert, R. & Henderson, J. M. (2010) CRISP: A computational model of fixation durations in scene viewing. *Psychological Review* 117(2):382–405. doi: 10.1037/a0018924. [ZS]
- O'Regan, J. K., Lévy-Schoen, A. & Jacobs, A. M. (1983) The effect of visibility on eye-movement parameters in reading. *Perception and Psychophysics* 34:457–64. doi: 10.3758/BF03203061. [aJH]
- Ohl, S., Brandt, S. A. & Kliegl, R. (2011) Secondary (micro-)saccades: The influence of primary saccade end point and target eccentricity on the process of post-saccadic fixation. *Vision Research* 51(23–24):2340–47. doi: 10.1016/j.visres.2011.09.005. [SO]
- Oliva, A. & Torralba, A. (2006) Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research* 155:23–36. doi: 10.1016/S0079-6123(06)55002-2. [AKH]
- Olivers, C. N. L. & Meeter, M. (2006) On the dissociation between compound and present/absent tasks in visual search: Intertrial priming is ambiguity driven. *Visual Cognition* 13(1):1–28. [aJH]
- Olivers, C. N. L., Meijer, F. & Theeuwes, J. (2006) Feature-based memory-driven attentional capture: Visual working memory content affects visual attention.

- Journal of Experimental Psychology: Human Perception and Performance* 32:1243–65. [KRC]
- Olivers, C. N. L. & Van der Helm, P. A. (1998) Symmetry and selective attention: A dissociation between effortless perception and visual search. *Perception and Psychophysics* 60:1101–16. doi: 10.3758/BF03206161. [aJH]
- Olivers, C. N., Peters, J., Houtkamp, R. & Roelfsema, P. R. (2011) Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences* 15(7):327–34. [RW]
- Or, C. C., Peterson, M. F. & Eckstein, M. P. (2015) Initial eye movements during face identification are optimal and similar across cultures. *Journal of Vision* 15 (13):12. [LH]
- Over, E. A. B., Hooge, I. T. C., Vlaskamp, B. N. S. & Erkelens, C. J. (2007) Coarse-to-fine eye movement strategy in visual search. *Vision Research* 47:2272–80. doi: 10.1016/j.visres.2007.05.002. [aJH, TM]
- Overvliet, K. E., Smeets, J. B. & Brenner, E. (2007) Haptic search with finger movements: Using more fingers does not necessarily reduce search times. *Experimental Brain Research* 182:427–34. [AP]
- Overvliet, K. E., Smeets, J. B. & Brenner, E. (2008) The use of proprioception and tactile information in haptic search. *Acta Psychologica* 129:83–90. [AP]
- Palmer, J., Ames, C. T. & Lindsey, D. T. (1993) Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance* 19(1):108–30. Available at: <http://doi.org/10.1037/0096-1523.19.1.108>. [AKr]
- Palmer, J., Verghese, P. & Pavel, M. (2000) The psychophysics of visual search. *Vision Research* 40:1227–68. doi: 10.1016/S0042-6989(99)00244-8. [arJH]
- Palmer, S. E. (1999) *Vision science: Photons to phenomenology*. MIT Press. [JC]
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A. & Morgan, M. (2001) Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience* 4(7):739–44. [RR]
- Pascual-Leone, A. & Hamilton, R. (2001) The metamodal organization of the brain. *Progress in Brain Research* 134:427–45. [AP]
- Pashler, H. (1987) Detecting conjunctions of color and form: Reassessing the serial search hypothesis. *Perception and Psychophysics* 41:191–201. doi: 10.3758/BF03208218. [aJH]
- Pasqualotto, A., Dumitru, M. L. & Myachikov, A. (2016) Multisensory integration: Brain, body, and world. *Frontiers in Psychology* 6:2046. [AP]
- Pasqualotto, A., Finucane, C. M. & Newell, F. N. (2005) Visual and haptic representations of scenes are updated with observer movement. *Experimental Brain Research* 166:481–88. [AP]
- Pasqualotto, A., Finucane, C. M. & Newell, F. N. (2013a) Ambient visual information confers a context-specific, long-term benefit on memory for haptic scenes. *Cognition* 128:363–79. [AP]
- Pasqualotto, A., Proulx, M. J. & Sereno, M. I. (2013b) Visual experience and the establishment of tactile face-maps in the brain. *Perception* 39:54. [AP]
- Pelli, D. G., Palomares, M. & Majaj, N. J. (2004) Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision* 4 (12):12. doi: 10.1167/4.12.12. [aJH]
- Peterson, M. A., Gillam, B. & Sedgwick, H. A., eds. (2006) *In the mind's eye: Julian Hochberg on the perception of pictures, films, and the world*. Oxford University Press. [JTE]
- Peterson, M. F. & Eckstein, M. P. (2012) Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences of the United States of America* 109(48):E3314–23. [LH]
- Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E. & McCarley, J. S. (2001) Visual search has memory. *Psychological Science* 12(4):287–92. [DEK]
- Phillips, S. (2014) Analogy, cognitive architecture and universal construction: A tale of two systematicities. *PLoS ONE* 9(2):e89152. [SP]
- Phillips, S., Takeda, Y. & Singh, A. (2012) Visual feature integration indicated by phase-locked frontal-parietal EEG signals. *PLoS ONE* 7(3):e32502. [SP]
- Phillips, S. & Wilson, W. H. (2010) Categorical compositionality: A category theory explanation for the systematicity of human cognition. *PLoS Computational Biology* 6(7):e1000858. [SP]
- Phillips, S. & Wilson, W. H. (2011) Categorical compositionality II: Universal constructions and a general theory of (quasi-)systematicity in human cognition. *PLoS Computational Biology* 7(8):e1002102. [SP]
- Phillips, S. & Wilson, W. H. (2012) Categorical compositionality III: F-(co)algebras and the systematicity of recursive capacities in human cognition. *PLoS ONE* 7 (4):e35028. [SP]
- Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W. H. C., Cohen, L. G., Guazzelli, M. & Haxby, J. V. (2004) Beyond sensory images: Object-based representation in the human ventral pathway. *Proceedings of the National Academy of Sciences of the United States of America* 101:5658–63. [AP]
- Pihko, E., Virtanen, A., Saarinen, V. M., Pannasch, S., Hirvankari, L., Tossavainen, T., Haapala, A. & Hari, R. (2011) Experiencing art: The influence of expertise and painting abstraction level. *Frontiers in Human Neuroscience* 5:94. [LH]
- Pirras, A. & Vickers, J. N. (2011) The effect of fixation transitions on quiet eye duration and performance in the soccer penalty kick: Instep versus inside kicks. *Cognitive Processing* 12:245–55. doi: 10.1007/s10339-011-0406-z. [JvdK, rJH]
- Plaisier, M., Kuling, I., Tiest, W. M. B. & Kappers, A. M. (2009) The role of item fixation in haptic search. In: *World Haptics 2009 – Third Joint EuroHaptics conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, Salt Lake City, UT, March 2009*, pp. 417–21. IEEE. doi: 10.1109/WHC.2009.4810798. [AP]
- Plaisier, M. A., Tiest, W. M. B. & Kappers, A. M. (2008) Haptic pop-out in a hand sweep. *Acta Psychologica* 128:368–77. [AP]
- Pöder, E. (2007) Effect of colour pop-out on the recognition of letters in crowding conditions. *Psychological Research* 71:641–45. doi: 10.1007/s00426-006-0053-7. [rJH]
- Pöder, E. (2008) Crowding with detection and coarse discrimination of simple visual features. *Journal of Vision* 8(4):24. doi: 10.1167/8.4.24. [aJH]
- Pöder, E. & Wagemans, J. (2007) Crowding with conjunctions of simple features. *Journal of Vision* 7(2):23. doi: 10.1167/7.2.23. [aJH, RR]
- Pomplun, M. (2007) Advancing area activation towards a general model of eye movements in visual search. In: *Integrated models of cognitive systems*, ed. W. D. Gray, pp. 120–31. Oxford University Press. [arJH]
- Pomplun, M., Garaas, T. W. & Carrasco, M. (2013) The effects of task difficulty on visual search strategy in virtual 3D displays. *Journal of Vision* 13(2013):1–22. doi: 10.1167/13.3.24. doi. [ZS]
- Pomplun, M., Reingold, E. M. & Shen, J. Y. (2003) Area activation: A computational model of saccadic selectivity in visual search. *Cognitive Science* 27:299–312. doi: 10.1016/S0364-0213(03)00003-X. [aJH, HJM]
- Preston, T. J., Guo, F., Das, K., Giesbrecht, B. & Eckstein, M. P. (2013) Neural representations of contextual guidance in visual search of real-world scenes. *Journal of Neuroscience* 33(18):7846–55. [LI]
- Prinz, W. (1977) Memory control of visual search. In: *Attention and performance VI*, ed. S. Dornic, pp. 441–62. Erlbaum. [WP, rJH]
- Prinz, W. (1979) Integration of information in visual search. *Quarterly Journal of Experimental Psychology* 31:287–304. [WP]
- Prinz, W. (1983) Asymmetrical control areas in continuous visual search. In: *Eye movements and psychological functions: International views*, ed. R. Groner, C. Menz, D. F. Fisher & R. A. Monty, pp. 85–110. Erlbaum. [WP]
- Prinz, W. (1986) Continuous selection. *Psychological Research* 48:231–38. [WP]
- Prinz, W. & Ataian, D. (1973) Two components and two stages in search performance: A case study in visual search. *Acta Psychologica* 37:255–77. [WP]
- Prinz, W. & Kehler, L. (1982) Recording detection distances in continuous visual search. In: *Cognition and eye movements*, ed. R. Groner & P. Fraisse, pp. 48–56. North-Holland. [WP]
- Prinz, W., Hartlich, S. & Lahmeyer, W. (1972) Nicht-visueller Transfer in einer visuellen Suchaufgabe. *Psychologische Forschung* 35:218–42. [WP]
- Prinz, W., Tweer, R. & Feige, R. (1974) Context control of search behavior: Evidence from a “hurdling”-technique. *Acta Psychologica* 38:73–80. [WP]
- Proulx, M. J., Parker, M. O., Tahir, Y. & Brennan, C. H. (2014) Parallel mechanisms for visual search in zebrafish. *PLoS ONE* 9:e111540. [AP]
- Pylyshyn, Z. W. & Storm, R. W. (1988) Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision* 3:179–97. doi: 10.1163/156856888X00122. [aJH]
- Rangelov, D., Müller, H. J. & Zehetleitner, M. (2013a) Visual search for feature singletons: Multiple mechanisms produce sequence effects in visual search. *Journal of Vision* 13:22. doi: 10.1167/13.3.22. [TT]
- Rangelov, D., Töllner, T., Müller, H. J. & Zehetleitner, M. (2013b) What are task-sets: A single, integrated representation or a collection of multiple control representations? *Frontiers in Human Neuroscience* 7:524. doi: 10.3389/fnhum.2013.00524. [TT]
- Rao, R., Zelinsky, G., Hayhoe, M. & Ballard, D. (2002) Eye movements in iconic visual search. *Vision Research* 42:1447–63. doi: 10.1016/S0042-6989(02)00040-8. [aJH]
- Rayner, K. (1986) Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology* 41:211–36. [JL]
- Rayner, K. (2009) Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology* 62:1457–506. Available at: <http://dx.doi.org/10.1080/17470210902516461>. [TJC]
- Rayner, K. & Duffy, S. A. (1986) Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition* 14:191–201. [TM]
- Rayner, K. & Fisher, D. L. (1987) Eye movements and the perceptual span during visual search. In: *Eye movements: From physiology to cognition*, ed. J. K. O’Regan & A. Levy-Schoen, pp. 293–302. Elsevier Science. [KRC]
- Reingold, E. M., Sheridan, H. (2011) Eye movements and visual expertise in chess and medicine. In: *Oxford handbook on eye movements*, ed. I. D. Gilchrist & S. Everling, pp. 528–50. Oxford University Press. [TJC]
- Reingold, E. M. & Claholt, M. G. (2014) Cognitive control of fixation duration in visual search: The role of extrafoveal processing. *Visual Cognition* 22(3):1–25. [TM]
- Reisberg, D. (2013) *Cognition: Exploring the science of the mind*, 5th edition. W. W. Norton & Company. [rJH]

- Reynolds, J. H. & Desimone, R. (1999) The role of neural mechanisms of attention in solving the binding problem. *Neuron* 24(1):19–29, 111–25. [LI]
- Riby, D. M., Brown, P. H., Jones, N. & Hanley, M. (2012) Brief report: Faces cause less distraction in autism. *Journal of Autism and Developmental Disorders* 42(4):634–39. [LH]
- Robertson, L. C. (2003) Binding, spatial attention and perceptual awareness. *Nature Reviews Neuroscience* 4(2):93–102. [LI]
- Roggeveen, A. B., Kingstone, A. & Enns, J. T. (2004) Influence of inter-item symmetry in visual search. *Spatial Vision* 17:443–64. [JTE]
- Rolfs, M. (2009) Microsaccades: Small steps on a long way. *Vision Research* 49(20):2415–41. doi: 10.1016/j.visres.2009.08.010. [SO]
- Rolfs, M. (2015) Attention in active vision. *Perception* 44(8–9):900–19. [SO]
- Rolfs, M., Jonikaitis, D., Deubel, H. & Cavanagh, P. (2011) Predictive remapping of attention across eye movements. *Nature Neuroscience* 14(2):252–56. doi: 10.1038/nn.2711. [SO]
- Rolfs, M., Kliegl, R. & Engbert, R. (2008) Toward a model of microsaccade generation: The case of microsaccadic inhibition. *Journal of Vision* 8(11):1–23. [SO]
- Rosenholtz, R. (2016) Capabilities and limitations of peripheral vision. *Annual Review of Vision Science* 2:437–57. [LH]
- Rosenholtz, R., Huang, J. & Ehinger, K. A. (2012a) Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology* 3(13):1–15. doi: 10.3389/fpsyg.2012.00013. [arJH, RR]
- Rosenholtz, R., Huang, J., Raj, A., Balas, B. J. & Ilie, L. (2012b) A summary statistic representation in peripheral vision explains visual search. *Journal of Vision* 12(4):14, 1–17. doi: 10.1167/12.4.14. [AL, RR, rJH]
- Rosenholtz, R., Li, Y. & Nakano, L. (2007) Measuring visual clutter. *Journal of Vision* 7(2):17, 1–22. doi: 10.1167/7.2.17. [RR]
- Rosenholtz, R. & Wijnntjes, M. (2014) Peripheral object recognition with informative natural context. *Journal of Vision* 14(10):214. [RR]
- Rubenstein, B. S. & Sagi, D. (1996) Preattentive texture segmentation: The role of line terminations, size, and filter wavelength. *Perception and Psychophysics* 58(4):489–509. [RR]
- Saenz, M., Buracas, G. T. & Boynton, G. M. (2002) Global effects of feature-based attention in human visual cortex. *Nature Neuroscience* 5:631–32. [ME]
- Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996) Statistical learning by 8-month-old infants. *Science* 274:1926–28. [RW]
- Sanders, A. F. (1970) Some aspects of the selective process in the functional visual field. *Ergonomics* 13:101–17. doi: 10.1080/00140137008931124. [aJH, KRC]
- Sann, C. & Streri, A. (2007) Perception of object shape and texture in human newborns: Evidence from cross-modal transfer tasks. *Developmental Science* 10:399–410. [AP]
- Savelsbergh, G. J. P., Williams, M. A., van der Kamp, J. & Ward, P. (2002) Visual search, anticipation and expertise in soccer goalkeepers. *Journal of Sports Sciences* 20:279–87. [jvdK]
- Schad, D. J. & Engbert, R. (2012) The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. *Visual Cognition* 20:391–421. [JL]
- Schapiro, A. C., Kustner, L. V. & Turk-Browne, N. B. (2012) Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology* 22:1622–27. [RW]
- Schmalzl, L., Palermo, R., Green, M., Brunson, R. & Coltheart, M. (2008) Training of familiar face recognition and visual scan paths for faces in a child with congenital prosopagnosia. *Cognitive Neuropsychology* 25(5):704–29. [DPC]
- Schneider, W. X. (2013) Selective visual processing across competition episodes: A theory of task-driven visual attention and working memory. *Philosophical Transactions of the Royal Society B* 368(1628):20130060. doi: 10.1098/rstb.2013.0060. [WP]
- Schneider, W. X., Einhäuser, W. & Horstmann, G. (2013) Attentional selection in visual perception, memory and action: A quest for cross-domain integration. *Philosophical Transactions of the Royal Society B* 368(1628):20130053. doi: 10.1098/rstb.2013.0053. [WP]
- Schuchard, R. A., Naseer, S. & de Castro, K. (1999) Characteristics of AMD patients receiving visual rehabilitation. *Journal of Rehabilitation Research and Development* 36(4):294–302. [DPC]
- Schumacher, E. H., Seymour, T. L., Glass, J. M., Fencsik, D., Lauber, E. J., Kieras, D. E. & Meyer, D. E. (2001) Virtually perfect time-sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psychological Science* 12:101–108. [DEK]
- Scialfa, C. T. & Joffe, K. M. (1998) Response times and eye movements in feature and conjunction search as a function of target eccentricity. *Perception and Psychophysics* 60:1067–82. doi: 10.3758/BF03211940. [aJH]
- Seiple, W., Rosen, R. B. & Garcia, P. M. T. (2013) Abnormal fixation in individuals with age-related macular degeneration when viewing an image of a face. *Optometry and Vision Science* 90(1):45–56. [DPC]
- Sekuler, A. B., Bennett, P. J. & Mamelak, M. (2000) Effects of ageing on the useful field of view. *Experimental Ageing Research* 26:103–20. doi: 10.1080/036107300243588. [aJH]
- Selfridge, O. (1959) Pandemonium: A paradigm for learning. In: *Symposium on the Mechanization of Thought Processes, National Physical Laboratory, United Kingdom, November 1958*, pp. 513–31. HM Stationery Office. [WP]
- Seymour, K., Clifford, C. W. G., Logothetis, N. K. & Bartels, A. (2009) The coding of color, motion, and their conjunction in the human visual cortex. *Current Biology* 19:177–83. doi: 10.1016/j.cub.2008.12.050. [rJH]
- Seymour, K., Clifford, C. W. G., Logothetis, N. K. & Bartels, A. (2010) Coding and binding of color and form in visual cortex. *Cerebral Cortex* 20:1946–54. doi: 10.1093/cercor/bhp265. [rJH]
- Shioiri, S. & Ikeda, M. (1989) Useful resolution for picture perception as a function of eccentricity. *Perception* 18:347–61. [JL]
- Shipp, S., Adams, D. L., Moutoussis, K. & Zeki, S. (2009) Feature binding in the feedback layers of area V2. *Cerebral Cortex* 19:2230–39. doi: 10.1093/cercor/bhn243. [rJH]
- Shomstein, S. S. & Yantis, S. (2002) Object-based attention: Sensory modulation or priority setting? *Perception and Psychophysics* 64(1):41–51. [KRC, WBUP]
- Simon, T. J. & Vaishnavi, S. (1996) Subitizing and counting depend on different attentional mechanisms: Evidence from visual enumeration in afterimages. *Perception and Psychophysics* 58:915–26. [DGW]
- Sirigu, A., Daprati, E., Pradat-Diehl, P., Franck, N. & Jeannerod, M. (1999) Perception of self-generated movement following left parietal lesion. *Brain* 122:1867–74. [JC]
- Smilek, D., Enns, J. T., Eastwood, J. D. & Merikle, P. M. (2006) Relax! Cognitive strategy influences visual search. *Visual Cognition* 14(4–8):543–64. [JTE]
- Smith, E. E. & Egeth, H. (1966) Effects of association value on perceptual search. *Journal of Experimental Psychology* 71:687–90. doi: 10.1037/h0023090. [aJH]
- Smith, N. D., Crabb, D. P. & Garway-Heath, D. F. (2011) An exploratory study of visual search performance in glaucoma. *Ophthalmic and Physiological Optics* 31(3):225–32. [DPC]
- Smith, N. D., Glen, F. C. & Crabb, D. P. (2012) Eye movements during visual search in patients with glaucoma. *BMC Ophthalmology* 12:45. [DPC]
- Song, C., Schwarzkopf, D. S., Kanai, R. & Rees, G. (2015) Neural population tuning links visual cortical anatomy to human visual perception. *Neuron* 85:641–56. doi: 10.1016/j.neuron.2014.12.041. [rJH]
- Soto, D. & Blanco, M. J. (2004) Spatial attention and object-based attention: A comparison within a single task. *Vision Research* 44:69–81. [WBUP]
- Sperlich, A., Schad, D. J. & Laubrock, J. (2015) When preview information starts to matter: Development of the perceptual span in German beginning readers. *Journal of Cognitive Psychology* 27:511–30. [JL]
- Sternberg, R. J. (2017) *Cognitive psychology*, seventh edition. Wadsworth Cengage. [rJH]
- Strasburger, H. (2005) Unfocused spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision* 5(11):1024–37. [RR]
- Sung, K. (2008) Serial and parallel attentive visual searches: Evidence from cumulative distribution functions of response times. *Journal of Experimental Psychology: Human Perception and Performance* 34:1372. [DRL]
- Sutton, R. S. & Barto, A. G. (1998) *Reinforcement learning: An Introduction*. The MIT Press. [JTE]
- Theeuwes, J., Godijn, R. & Pratt, J. (2004) A new estimate of attentional dwell time. *Psychonomic Bulletin and Review* 11:60–64. doi: 10.3758/BF03206461. [arJH]
- Theeuwes, J., Kramer, A. F., Hahn, S. & Irwin, D. E. (1998) Our eyes do not always go where we want them to go: Capture of eyes by new objects. *Psychological Science* 9:379–85. [SVdS]
- Theeuwes, J., Mathôt, S. & Grainger, J. (2013) Exogenous object-centered attention. *Attention Perception, and Psychophysics* 75:812–18. doi: 10.3758/s13414-013-0459-4. [SVdS, rJH]
- Theeuwes, J., Mathôt, S. & Kingstone, A. (2010) Object-based eye movements: The eyes prefer to stay within the same object. *Attention Perception, and Psychophysics* 72(3):12–21. doi: 10.3758/APP.72.3.597. [SVdS, rJH]
- Theeuwes, J., Mathôt, S. & Grainger, J. (2013) Exogenous object-centered attention. *Attention, Perception, and Psychophysics* 75:812–18. [SVdS]
- Thibaut, M., Delerue, C., Boucart, M. & Tran, T. H. C. (2015) Visual exploration of objects and scenes in patients with age-related macular degeneration. *Journal Français d'Ophtalmologie* 39:82–89. [DPC]
- Thompson, E. R., Iyer, N., Simpson, B. D., Wakefield, G. H., Kieras, D. E. & Brungart, D. S. (2015) Enhancing listener strategies using a payoff matrix in speech-on-speech masking experiments. *Journal of the Acoustical Society of America* 138(3):1297–304. [DEK]
- Thorpe, S., Fize, D. & Marlot, C. (1996) Speed of processing in the visual system. *Nature* 381:520–22. doi: 10.1038/381520a0. [aJH]
- Töllner, T., Conci, M. & Müller, H. J. (2015a) Predictive distractor context facilitates attentional selection of high, but not intermediate and low, salience targets. *Human Brain Mapping* 36(3):935–44. doi: 10.1002/hbm.22677. [TT]
- Töllner, T., Conci, M., Rusch, T. & Müller, H. J. (2013) Selective manipulation of target identification demands in visual search: The role of stimulus contrast in CDA activations. *Journal of Vision* 13(3):23, 1–13. doi: 10.1167/13.3.23. [TT]
- Töllner, T., Eschmann, K., Rusch, T. & Müller, H. J. (2014) Contralateral delay activity reveals dimension-based attentional orienting to locations in visual

- working memory. *Attention Perception, and Psychophysics* 76(3):655–62. doi: 10.3758/s13414-014-0636-0. [TT]
- Töllner, T., Mink, M. & Müller, H. J. (2015b) Searching for targets in visual working memory: Investigating a “dimensional feature bundle” (DFB) model. *Annals of the New York Academy of Sciences* 1339(1):32–44. doi: 10.1111/nyas.12703. [TT]
- Töllner, T., Müller, H. J. & Zehetleitner, M. (2012a) Top-down dimensional weight set determines the capture of visual attention: Evidence from the PCN component. *Cerebral Cortex* 22(7):1554–63. doi: 10.1093/cercor/bhr231. [TT]
- Töllner, T., Rangelov, D. & Müller, H. J. (2012b) How the speed of motor-response decisions, but not focal-attentional selection, differs as a function of task set and target prevalence. *Proceedings of the National Academy of Sciences of the United States of America* 109:E1990–99. doi: 10.1073/pnas.1206382109. [aJH, HJM, TT]
- Töllner, T., Zehetleitner, M., Gramann, K. & Müller, H. J. (2011) Stimulus saliency modulates pre-attentive processing speed in human visual cortex. *PLoS ONE* 6(1):e16276. doi: 10.1371/journal.pone.0016276. [TT]
- Tomonaga, M. (2007) Visual search for orientation of faces by a chimpanzee (*Pan troglodytes*): Face-specific upright superiority and the role of facial configural properties. *Primates* 48:1–12. [AP]
- Torrallba, A., Oliva, A., Castelhan, M. S. & Henderson, J. M. (2006) Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review* 113(4):766–86. [LH, LI]
- Townsend, J. T. (1971) A note on the identifiability of parallel and serial processes. *Perception and Psychophysics* 10:161–63. [DRL]
- Townsend, J. T. (1972) Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology* 25:168–99. [DRL]
- Townsend, J. T. (1990) Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science* 1:46–54. [DRL]
- Townsend, J. T. & Ashby, F. G. (1983) *The stochastic modeling of elementary psychological processes*. Cambridge University Press. [DRL]
- Townsend, J. T. & Nozawa, G. (1995) Spatio-temporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology* 39:321–40. [DRL]
- Townsend, J. T. & Wenger, M. J. (2004) A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review* 111:1003–1035. [DRL]
- Trappenberg, T. P., Dorris, M. C., Munoz, D. P. & Klein, R. M. (2001) A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitive Neuroscience* 13(2):256–71. [SVdS]
- Treisman, A. (1982) Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance* 8:194–214. [aJH, AP]
- Treisman, A. (1991) Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance* 17:652–76. doi: 10.1037/0096-1523.17.3.652. [aJH]
- Treisman, A. (1996) The binding problem. *Current Opinion in Neurobiology* 6:171–78. [JMW]
- Treisman, A. (2006) How the deployment of attention determines what we see. *Visual Cognition* 14:411–43. doi: 10.1080/13506280500195250. [aJH]
- Treisman, A. & Schmidt, H. (1982) Illusory conjunctions in the perception of objects. *Cognitive Psychology* 14:107–41. doi: 10.1016/0010-0285(82)90006-8. [aJH]
- Treisman, A. & Gormican, S. (1988) Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review* 95:15–48. [KRC, rJH]
- Treisman, A. & Sato, S. (1990) Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance* 16:459–78. doi: 10.1037/0096-1523.16.3.459. [aJH]
- Treisman, A. M. & Gelade, G. (1980) A feature-integration theory of attention. *Cognitive Psychology* 12:97–136. doi: 10.1016/0010-0285(80)90005-5. [arJH, AKh, JvdK]
- Trick, L. M. & Pylyshyn, Z. W. (1994) Why are small and large numbers enumerated differently: A limited capacity preattentive stage in vision. *Psychological Review* 101:80–102. [DGW]
- Tsal, Y. & Benoni, H. (2010) Diluting the burden of load: Perceptual load effects are simply dilution effects. *Journal of Experimental Psychology: Human Perception and Performance* 36:1645–56. [KRC]
- Tseng, P., Carmi, R., Cameron, I. G. M., Munoz, D. P. & Itti, L. (2009) Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision* 9(7):1–16. [TM]
- Tsotsos, J. K. (1990) Analyzing vision at the complexity level. *Brain and Behavioral Sciences* 13(3):423–69. [JMW]
- Turk-Browne, N. B., Jungé, J. A. & Scholl, B. J. (2005) The automaticity of visual statistical learning. *Journal of Experimental Psychology: General* 134:552–64. [RW]
- Uesaki, M. & Ashida, H. (2015) Optic-flow selective cortical sensory regions associated with self-reported states of vection. *Frontiers in Psychology* 6:775. [AP]
- Unema, P. J. a., Pannasch, S., Joos, M. & Velichkovsky, B. M. (2005) Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition* 12(3):473–94. doi: 10.1080/1350628044000409. [LH, ZS]
- van Asselen, M., Sampaio, J., Pina, A. & Castelo-Branco, M. (2011) Object based implicit contextual learning: A study of eye movements. *Attention, Perception, and Psychophysics* 73(2):297–302. doi: 10.3758/s13414-010-0047-9. [ZS]
- van den Berg, R., Johnson, A., Martinez Anton, A., Schepers, A. L. & Cornelissen, F. W. (2012) Comparing crowding in human and ideal observers. *Journal of Vision* 12(6):13 doi: 10.1167/12.6.13. [RR]
- van Doorn, H., van der Kamp, J., de Wit, M. & Savelsbergh, G. J. P. (2009) Another look at the Müller-Lyer illusion: Different gaze patterns in vision for action and perception. *Neuropsychologia* 47:804–12. [JvdK]
- van Gelder, T. (1990) Compositionality: A connectionist variation on a classical theme. *Cognitive Science* 14:355–84. [SP]
- Van Polanen, V., Tiest, W. M. B. & Kappers, A. M. (2011) Movement strategies in a haptic search task. In: *World Haptics Conference, 2011 IEEE*, 275–80. [AP]
- van Zoest, W., Giesbrecht, B., Enns, J. T. & Kingstone, A. F. (2006) Inter-item symmetry influences visual search. *Psychological Science* 17:535–42. [JTE]
- van Zoest, W., Lleras, A., Kingstone, A. F. & Enns, J. T. (2007) In sight, out of mind: The role of eye movements in the rapid resumption of visual search. *Perception and Psychophysics* 69:1204–17. [JTE]
- Vatterott, D. B. & Vecera, S. P. (2015) The attentional window configures to object and surface boundaries. *Visual Cognition* 23(5):561–76. Available at: <http://doi.org/10.1080/13506285.2015.1054454>. [AKr]
- Verghese, P. (2001) Visual search and attention: A signal detection approach. *Neuron* 31:523–35. doi: 10.1016/S0896-6273(01)00392-0. [arJH]
- Virsu, V. & Rovamo, J. (1979) Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental Brain Research* 37:475–94. [DEK]
- Vö, M. L. H. & Wolfe, J. M. (2012) When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *Journal of Experimental Psychology: Human Perception and Performance* 38:23–41. doi: 10.1037/a0024147. [aJH]
- Vogel, E. K. & Machizawa, M. G. (2004) Neural activity predicts individual differences in visual working memory capacity. *Nature* 428:748–51. doi: 10.1038/nature02447. [TT]
- Volkman, F. C., Riggs, L. A., White, K. D. & Moore, R. K. (1978) Contrast sensitivity during saccadic eye movements. *Vision Research* 18:1193–99. doi: 10.1016/0042-6989(78)90104-9. [aJH]
- Wang, D., Kristjánsson, Á. & Nakayama, K. (2005) Efficient visual search without top-down or bottom-up guidance. *Perception and Psychophysics* 67:239–53. [AKr]
- Wang, S., Jiang, M., Duchesne, X. M., Laugeson, E. A., Kennedy, D. P., Adolphs, R. & Zhao, Q. (2015) Atypical visual saliency in Autism spectrum disorder quantified through model-based eye tracking. *Neuron* 88(3):604–16. [LH]
- Wannig, A., Stanisor, L. & Roelfsema, P. R. (2011) Automatic spread of attention response modulation along Gestalt criteria in primary visual cortex. *Nature Neuroscience* 14(10):1243–44. [WBUP]
- Ward, J. (2015) *The student's guide to cognitive neuroscience*, 3rd edition. Psychology Press. [rJH]
- Watson, D. G. & Humphreys, G. W. (1997) Visual marking: Prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychological Review* 104(1):90–122. [DGW]
- Watson, D. G. & Humphreys, G. W. (1998) Visual marking of moving objects: A role for top-down feature-based inhibition in selection. *Journal of Experimental Psychology: Human Perception and Performance* 24:946–62. [DGW]
- Watson, D. G. & Humphreys, G. W. (2000) Visual marking: Evidence for inhibition using a probe-dot detection paradigm. *Perception and Psychophysics* 62:471–81. doi: 10.3758/BF03212099. [aJH]
- Watson, D. G., Maylor, E. A. & Bruce, L. A. M. (2005) Search, enumeration, and aging: Eye movement requirements cause age-equivalent performance in enumeration but not in search tasks. *Psychology and Aging* 20:226–40. [DGW]
- Watson, D. G., Maylor, E. A. & Bruce, L. A. M. (2007) The role of eye movements in subitizing and counting. *Journal of Experimental Psychology: Human Perception and Performance* 33:1389–99. [DGW]
- Watson, M. R., Brennan, A. A., Kingstone, A. & Enns, J. T. (2010) Looking versus seeing: Strategies alter eye movements during visual search. *Psychonomic Bulletin and Review* 17(4):543–49. [JTE]
- Weidner, R. & Müller, H. J. (2013) Dimensional weighting in cross-dimensional singleton conjunction search. *Journal of Vision* 13(3):25. doi: 10.1167/13.3.25. [HJM]
- Westheimer, G. (1982) The spatial grain of the perifoveal visual field. *Vision Research* 22:157–62. doi: 10.1016/0042-6989(82)90177-8. [aJH]
- Wiegand, I., Töllner, T., Habekost, T., Dyhrholm, M., Müller, H. J. & Finke, K. (2014) Distinct neural markers of TVA-based visual processing speed and short-term storage capacity parameters. *Cerebral Cortex* 24(8):1967–78. doi: 10.1093/cercor/bht071. [TT]

- Williams, D. E., Reingold, E. M., Moscovitch, M. & Behrmann, M. (1997) Patterns of eye movements during parallel and serial visual search tasks. *Canadian Journal of Experimental Psychology* 51:151–64. doi: 10.1037/1196-1961.51.2.151. [rJH]
- Williams, L. J. (1982) Cognitive load and the functional field of view. *Human Factors* 24:683–92. [rJH]
- Wilson, D. E., Muroi, M. & MacLeod, C. M. (2011) Dilution, not load, affects distractor processing. *Journal of Experimental Psychology: Human Perception and Performance* 37:319–35. [KRC]
- Wolfe, J. M. (1992) “Effortless” texture segmentation and “parallel” visual search are not the same thing. *Vision Research* 32:757–63. doi: 10.1016/0042-6989(92)90190-T. [aJH]
- Wolfe, J. M. (1994) Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review* 1(2):202–38. [aJH, AL]
- Wolfe, J. M. (1998a) Visual search. In: *Attention*, ed. H. Pashler, pp. 13–73. University College London Press. [aJH]
- Wolfe, J. M. (1998b) What can 1 million trials tell us about visual search? *Psychological Science* 9:33–39. doi: 10.1111/1467-9280.00006. [arJH, DRL, AP]
- Wolfe, J. M. (2003) Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences* 7:70–76. doi: 10.1016/S1364-6613(02)00024-4. [aJH]
- Wolfe, J. M. (2007) Guided Search 4.0: Current Progress with a model of visual search. In: *Integrated models of cognitive systems*, ed. W. Gray, pp. 99–119. Oxford University Press. [aJH]
- Wolfe, J. M. (2010) Guided Search 4.0: A guided search model that does not require memory for rejected distractors. *Journal of Vision* 1(3):349–49. doi: 10.1167/1.3.349. [WBUP]
- Wolfe, J. M., Alvarez, G. A. & Horowitz, T. S. (2000) Attention is fast but volition is slow. *Nature* 406:691. doi: 10.1038/35021132. [rJH]
- Wolfe, J. M., Alvarez, G. A., Rosenholtz, R. E., Kuzmina, Y. I. & Sherman, A. M. (2011a) Visual search for arbitrary objects in real scenes. *Attention, Perception, and Psychophysics* 73:1650–71. doi: 10.3758/s13414-011-0153-3. [aJH, AL, RR]
- Wolfe, J. M. & Bennett, S. C. (1997) Preattentive object files: Shapeless bundles of basic features. *Vision Research* 37(1):25–43. [JMW]
- Wolfe, J. M., Brunelli, D. N., Rubinstein, J. & Horowitz, T. S. (2013) Prevalence effects in newly trained airport checkpoint screeners: Trained observers miss rare targets, too. *Journal of Vision* 13(3):33. doi: 10.1167/13.3.33. [aJH]
- Wolfe, J. M., Cave, K. R. & Franzel, S. L. (1989) Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance* 15(3):419–33. [aJH, SP]
- Wolfe, J. M. & Gancarz, G. (1996) Guided Search 3.0: A model of visual search catches up with Jay Enoch 40 years later. In: *Basic and clinical applications of vision science*, ed. V. Lakshminarayanan, pp. 189–92. Kluwer Academic. [aJH]
- Wolfe, J. M. & Horowitz, T. S. (2004) What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience* 5:495–501. doi: 10.1038/nrn1411. [arJH]
- Wolfe, J. M., Horowitz, T. S. & Kenner, N. M. (2005) Rare items often missed in visual searches. *Nature* 435(7041):439–40. doi: 10.1038/435439a. [aJH, TM]
- Wolfe, J. M., Horowitz, T. S. & Palmer, E. M. (2010a) RT-distributions constrain models of visual search. *Vision Research* 50:1304–11. doi: 10.1016/j.visres.2009.11.002. [aJH]
- Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzky, R. L., Lederman, S. J. & Merfeld, D. M. (2015) *Sensation & perception*, 4th edition. Sinauer Associates. [rJH]
- Wolfe, J. M., O’Neill, P. & Bennett, S. C. (1998) Why are there eccentricity effects in visual search? Visual and attentional hypotheses. *Perception and Psychophysics* 60:140–56. doi: 10.3758/BF03211924. [aJH]
- Wolfe, J. M., Palmer, E. M. & Horowitz, T. S. (2010b) Reaction time distributions constrain models of visual search. *Vision Research* 50:1304–11. [DRL, RM, AP]
- Wolfe, J. M., Võ, M. L. H., Evans, K. K. & Greene, M. R. (2011b) Visual search in scenes involves selective and non-selective pathways. *Trends in Cognitive Sciences* 15:77–84. doi: 10.1016/j.tics.2010.12.001. [aJH]
- Woodman, G. F. & Luck, S. J. (1999) Electrophysiological measurement of rapid shifts of attention during visual search. *Nature* 400:867–69. doi: 10.1038/23698. [KRC, rJH]
- Woodman, G. F. & Luck, S. J. (2003) Serial deployment of attention during visual search. *Journal of Experimental Psychology: Human Perception and Performance* 29:121–38. doi: 10.1037/0096-1523.29.1.121. [rJH]
- Woods, D. L., Courchesne, E., Hillyard, S. A. & Galambos, R. (1980) Recovery cycles of event-related potentials in multiple detection tasks. *Electroencephalography and Clinical Neurophysiology* 50:335–47. doi: 10.1016/0013-4694(80)90001-2. [TT]
- Wu, R., Gopnik, A., Richardson, D. C. & Kirkham, N. Z. (2011) Infants learn about objects from statistics and people. *Developmental Psychology* 47(5):1220–29. [RW]
- Wu, R., Nako, R., Band, J., Pizzuto, J., Shadravan, Y., Scerif, G. & Aslin, R. N. (2015) Rapid selection of non-native stimuli despite perceptual narrowing. *Journal of Cognitive Neuroscience* 27(11):2299–307. [RW]
- Wu, R., Pruitt, Z., Runkle, M., Scerif, G. & Aslin, R. N. (2016) A neural signature of rapid category-based target selection as a function of intra-item perceptual similarity despite inter-item dissimilarity. *Attention Perception and Psychophysics* 78(3):749–76. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26732265>. [RW]
- Wu, R., Pruitt, Z., Zinszer, B. & Cheung, O. (2017) Increased experience amplifies the activation of task-irrelevant category representations. *Attention, Perception, and Psychophysics* 79(2):522–32. [RW]
- Wu, R., Scerif, G., Aslin, R. N., Smith, T. J., Nako, R. & Eimer, M. (2013) Searching for something familiar or novel: Top-down attentional selection of specific items or object categories. *Journal of Cognitive Neuroscience* 25(5):719–29. [RW]
- Wurtz, R. H. (2008) Neuronal mechanisms of visual stability. *Vision Research* 48(20):2070–89. doi: 10.1016/j.visres.2008.03.021. [SO]
- Xiaodi, H., Harel, J. & Koch, C. (2012) Image signature: Highlighting sparse salient regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(1):194–201. doi: 10.1109/TPAMI.2011.146. [AKh]
- Yarbus, A. L. (1967) *Eye movements and vision*. Plenum Press. [LH]
- Young, A. H. & Hulleman, J. (2013) Eye movements reveal how task difficulty moulds visual search. *Journal of Experimental Psychology: Human Perception and Performance* 39:168–90. [arJH, KRC, AP, TT]
- Yu, R. & Zhao, J. (2015) The persistence of attentional bias to regularities in a changing environment. *Attention, Perception, and Psychophysics* 77:2217–28. [RW]
- Yuval-Greenberg, S., Merriam, E. P. & Heeger, D. J. (2014) Spontaneous microsaccades reflect shifts in covert attention. *Journal of Neuroscience* 34(41):13693–700. doi: 10.1523/JNEUROSCI.0582-14.2014. [SO]
- Zang, X., Jia, L., Müller, H. J. & Shi, Z. (2015) Invariant spatial context is learned but not retrieved in gaze-contingent limited-viewing search. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41(3):807–19. doi: 10.1037/xlm0000060. [ZS]
- Zelinsky, G. J. (1996) Using eye saccades to assess the selectivity of search movements. *Vision Research* 36:2177–21. [aJH, AP]
- Zelinsky, G. J. (2008) A theory of eye movements during target acquisition. *Psychological Review* 115:787–835. doi: 10.1037/a0013118. [aJH, AL, HJM, RR]
- Zelinsky, G. J. (2012) TAM: Explaining off-object fixations and central fixation tendencies as effects of population averaging during search. *Visual Cognition* 20:515–45. doi: 10.1080/13506285.2012.666577. [aJH]
- Zelinsky, G. J., Adeli, H., Peng, Y. & Samaras, D. (2013) Modelling eye movements in a categorical search task. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 368(1628):20130058. [SP, aJH]
- Zelinsky, G. J. & Sheinberg, D. L. (1995) Why some search tasks take longer than others: Using eye movements to redefine reaction times. In: *Eye movement research: Mechanism, processes and applications*, ed. J. M. Findlay, R. Walker & R. W. Kentridge, pp. 325–36. Elsevier Science. doi: 10.1016/S0926-907X(05)80028-5. [arJH]
- Zelinsky, G. J. & Sheinberg, D. L. (1997) Eye movements during parallel-serial visual search. *Journal of Experimental Psychology: Human Perception and Performance* 23:244–62. doi: 10.1037/0096-1523.23.1.244. [arJH]
- Zelinsky, G. J., Peng, Y., Berg, A. C. & Samaras, D. (2013) Modeling guidance and recognition in categorical search: Bridging human and computer object detection. *Journal of Vision* 13(3):30. 1–20. doi: 10.1167/13.3.30. [rJH]
- Zelinsky, G. J., Rao, R. P. N., Hayhoe, M. M. & Ballard, D. H. (1997) Eye movements reveal the spatiotemporal characteristics of visual search. *Psychological Science* 8:448–53. doi: 10.1111/j.1467-9280.1997.tb00459.x. [rJH]
- Zhang, X., Huang, J., Yigit-Elliott, S. & Rosenholtz, R. (2015) Cube search, revisited. *Journal of Vision* 15(3):9. 1–18. doi: 10.1167/15.3.9. [RR, rJH]
- Zhang, Y. & Hornof, A. J. (2014) Understanding multitasking through parallelized strategy exploration and individualized cognitive modeling. In: *CHI ’14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3885–94. ACM. [DEK]
- Zhang, Y., Meyers, E. M., Bichot, N. P., Serre, T., Poggio, T. A. & Desimone, R. (2011) Object decoding with attention in inferior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America* 108(21):8850–55. [LI]
- Zhao, J., Al-Aidroos, N. & Turk-Browne, N. B. (2013) Attention is spontaneously biased toward regularities. *Psychological Science* 24:667–77. [RW]
- Zhao, J. & Luo, Y. (2014) Statistical regularities alter the spatial scale of attention. *Journal of Vision* 14(10):11. [RW]
- Zhao, J., Ngo, N., McKendrick, R. & Turk-Browne, N. B. (2011) Mutual interference between statistical summary perception and statistical learning. *Psychological Science* 22:1212–19. [RW]
- Zhao, J. & Yu, R. (2016) Statistical regularities reduce perceived numerosity. *Cognition* 146:217–22. [RW]
- Zou, H., Müller, H. J. & Shi, Z. (2012) Non-spatial sounds regulate eye movements and enhance visual search. *Journal of Vision* 12(5):2. 1–18. doi: 10.1167/12.5.2. [ZS]
- Zupan, Z., Watson, D. G. & Blagrove, E. (2015) Inhibition in time-based visual selection: Strategic or by default? *Journal of Experimental Psychology: Human Perception and Performance* 41:1442–61. [DGW]