Vision Research 51 (2011) 1526–1537

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Review

Situating visual search

Ken Nakayama[a,*], Paolo Martini[b,1]

[a] Department of Psychology, Harvard University, 33 Kirkland St. Cambridge, MA 02138, USA
[b] Department of Psychology, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

ARTICLE INFO

Article history:
Received 27 July 2010
Received in revised form 1 September 2010
Available online 15 September 2010

Keywords:
Object recognition
Attention
Reinforcement learning

ABSTRACT

Visual search attracted great interest because its ease under certain circumstances seemed to provide a way to understand how properties of early visual cortical areas could explain complex perception without resorting to higher order psychological or neurophysiological mechanisms. Furthermore, there was the hope that properties of visual search itself might even reveal new cortical features or dimensions. The shortcomings of this perspective suggest that we abandon fixed canonical elementary particles of vision as well as a corresponding simple to complex cognitive architecture for vision. Instead recent research has suggested a different organization of the visual brain with putative high level processing occurring very rapidly and often unconsciously. Given this outlook, we reconsider visual search under the broad category of recognition tasks, each having different trade-offs for computational resources, between detail and scope. We conclude noting recent trends showing how visual search is relevant to a wider range of issues in cognitive science, in particular to memory, decision making, and reward.

1. Introduction

During WWII the US Navy recognized a pressing need to develop systematic ways of locating enemy ships or it's own lost or missing personnel at sea. So it was that a group of mathematicians were tasked to develop, in secret, a Theory of Search. Those efforts were eventually declassified and published in a series of three articles, and eventually a book, by Bernard Koopman in the 1950s (Koopman, 1956a, 1956b, 1957, 1980). Koopman's articles and book are singular in one particular aspect: they contain no references, since, as their Author concluded, none existed then. Having no empirical evidence, Koopman theoretically derived and discussed from first principles many aspects of search behavior that would be investigated in the ensuing decades: the visual lobe, the form of the detection function, the optimal distribution of search effort, criteria for termination, etc. The direct heirs of such tradition can be found in the contemporary scientific domains of Operations Research, Ergonomics and Decision Theory.

As is often the case in the history of science, exact origins can be difficult to trace. Fifty years prior to Koopman, and well over 100 years ago, the famous British zoologist Edward Poulton, and early advocate of Darwinian theory, had his own theoretical speculations about visual search, this time in the context of animals more effectively eluding predators. Aside from considering the issue of camouflage where some animals are well concealed by taking on the color and texture of their environments, he also considered the issue of polymorphisms. In his book "The colors of animals" he noted that within a species, a variety of colors can exist. For example, the larvae of emerald moths came in green and brown. Why is this so? He wrote:

> " … Both forms are common, and therefore it is certain that both must be advantageous to the species, or one of them would quickly disappear. I believe that it is a benefit to the species that some of its larvae should resemble brown and others green catkins, instead of all of hem resembling either brown or green. In the former case the foes have a wider range of objects for which they may mistake the larvae, and the *search* must occupy more time, for equivalent results, than in the case of other species which are not dimorphic" (Poulton, 1890, p. 46).

To this day, both operations researchers and biologists remain interested in visual search. Added to this have been the increasingly large number of studies of visual search by psychologists, vision researchers and neuroscientists. Visual search paradigms are one of the most represented in journals and textbooks. The interest is so great that even a cursory look at the literature can be overwhelming. In the ISI database, there are over 2200 articles with the specific phrase "visual search" in the title.

Why has this become such a popular subject and what does it portend for those of us who have devoted their lives in and around the visual system and are trying to understand it? In contrast to many other vision laboratory paradigms, it has some clear advantages. First, it deals with supra-threshold vision such that the

[*] Corresponding author. Fax: +1 617 495 3764.
E-mail addresses: ken@wjh.harvard.edu (K. Nakayama), p.martini@warwick.ac.uk (P. Martini).
[1] Fax: +44 2476 524 225.

0042-6989/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2010.09.003

targets searched for are plainly visible. This links the phenomenon to something we do in our daily lives. And of course, visual search displays with its multiple items can be varied in simple ways, numbers, density, contrast, homogeneity, extent. Additionally, the items can be more or less anything – letters, Gabor patches, faces, abstract shapes, tools, such that it would seem that any level of vision can be probed. And finally, the trials are simple and can be repeated so as to get reliable measures. Finally, it can be used on young and old, infirm and in experimental animals.

But all of these reasons alone would not elicit great interest. What was needed was some thing more. We can make lots of systematic measurements but eventually we need to think we know why. Several scientific developments (in addition to practical ones) were at work to keep interest high.

Advances in single unit neurophysiology started around mid-century, just before the inception of this journal and they provided a dominant perspective for vision research for many decades. Early results on the frog retina by Barlow (1953) were followed quickly by "What the frog's eye tells the frog's brain" by Lettvin, Maturana, McCulloch, and Pitts (1959). These were complemented by the seminal work of Hubel and Wiesel (1959), and Hubel and Wiesel (1968) who systematically studied the receptive field characteristics of neurons in the cat and monkey visual system, discovering neurons specifically responsive to oriented edges and bars, later showing responses to specific colors and motion directions. Following these revolutionary discoveries, it became apparent to most everybody, that there exist retinotopic maps, each containing neurons which could code basic elementary features of the visual image. Reinforcing and finally canonizing these views, was the recognition of Hubel and Wiesel's work with the Nobel Prize in 1981.

Feeling the need for a more quantitative approach, a different breed of researchers were able to formalize cortical neurons as coding spatial frequency and orientation, stating that the earliest stages of cortex consisted of Gabor filters which had restricted sensitivity both to position and spatial frequency (Campbell, Cooper, & Enroth-Cugell, 1969; De Valois, Albrecht & Thorell, 1982). The results were both startling and significant for several reasons. First, they indicated that the spatial receptive fields described by Hubel and Wiesel were too coarsely characterized, that the tuning characteristics of cortical neurons required that the receptive field's of striate cortex neurons have more regions (excitatory and inhibitory) to account for the relatively narrow spatial frequency tuning. This Fourier approach was clearly at odds with the studies mentioned above, since the former approach described "features" (bars, edges) that for the most part could be "seen" in most natural and laboratory created images. In contrast, people could not "see" Gabor patches embedded within natural or most laboratory images. David Marr in his historic synthesis of vision, attempted to reconcile these different views by setting up detectors that could recover edges from these Gabor-like outputs (Marr, 1982; Marr & Hildreth, 1980).

Despite the doubts raised by spatial frequency advocates, by the 1970s many would at least tacitly assume that the early visual system analyzed the visual world in terms of features or visual dimensions, color, orientation, motion, depth, and scale although devotees of spatial frequency would stick to their own formulation. While Hubel and Wiesel were initially reluctant to suggest specific ways in which the receptive field properties of cells might be related to visual perception, other researchers saw the properties of these cells as a very powerful way to understand vision. As such, there was a great flourishing of human psychophysical studies confirming the existence of multiple spatial frequency channels (Blakemore & Campbell, 1969) as well as exploring many different topics through after-effects, such as the McCullough effect, the tilt after-effect, the motion after-effect and other related phenomena. All of these studies were consistent with the idea that at some locus in the visual system, probably in V1 and closely beyond, there were detectors of various sorts, tiling the visual field.

2. Visual search and brain architecture 1.0

With these important trends as background, two charismatic researchers supplied the kind of framework to put studies of visual search into the limelight: Bela Julesz and Anne Treisman. Bela Julesz was born in Hungary, a refugee who escaped in the wake of the 1956 Hungarian uprising and who in 1960 dazzled the psychological community with his random dot stereograms, the first truly exciting application of computers to psychology (Julesz, 1964). Anne Treisman, undergraduate student of modern languages at the University of Cambridge, switched to psychology and made an early name for herself in auditory attention, when visual attention was not yet a proper field.

While starting from very different premises and with different goals, each wanted to study and isolate putative elementary aspects of vision. Julesz's ambition was to find elementary particles of vision much as the physical sciences had found atoms, molecules. Because these putative "textons" were somehow fundamental visual elements, they would be apprehended very quickly without scrutiny and their registration in the nervous system would be governed by very simple rules, by changes in density of elements, terminators, oriented lines, closure. Julesz, however, did not explicitly think about these units in terms of the task of visual search but in terms of something closely related, "effortless" perception, using texture segregation as a way to determine whether a given texture element was a texton, an elementary unit (Julesz, 1981).

It was Treisman's Feature Integration Theory (FIT) that had the greatest influence in elevating the status of visual search because this paradigm was the main source of data for her theory and it was to be the proving ground for it to be tested more fully. At the core of the theory was the idea of distinct feature maps, two-dimensional arrays of detectors any one of which could be activated in parallel. There were feature maps say for color (red, green, blue, etc.), basic shapes (letters, geometric figures), and other properties (Treisman & Gelade, 1980). Visual search could occur in two modes, pre-attentive (not requiring attention) and attentive. In the pre-attentive mode, the activity of a singleton in a feature map, say a green square amongst a field of red squares would effortlessly "popout" by virtue of being a sole locus of activity in the green primitive feature map.

The operational definition of popout was boldly stated. Popout in a given search array occurs when performance (in terms of reaction time) remains constant, no matter how many distractors. According to the theory, this occurs because popout is mediated by the unique occurrence of a feature in a retinotopic map, all of this occurring in parallel. Thus there would be no cost in adding more distractors. Many singletons placed amongst distractors showed this characteristic function, thus the green item amongst red ones, X's among Os, high spatial frequency Gabor's amongst low frequency ones (Fig. 1a, b, and c respectively). Treisman claimed that these flat "popout" functions were a diagnostic characteristic of parallel search and hence could point to the existence of retinotopic maps of a given feature or dimension.

Contrasted with this, were searches for conjunctions of features, where no such parallel scheme of single feature maps existed and so conjunctive search could not be handled by the parallel operation of single feature maps by themselves. (see Fig. 1d). Under these circumstances, pre-attentive vision cannot occur. FIT postulated that attention was needed to bind features at some other more central locus, a master map of locations where attention

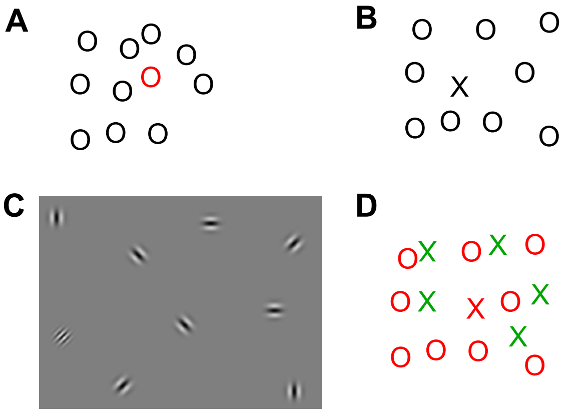

Fig. 1. Search for a single target varying in only one dimension (a,b,c) vs searching for the single target defined by the conjunction of two dimensions (d).

would have the role of binding features. The slope of the serial search function or search rate (milliseconds per item) provided the amount of time it took focal attention to move from one item to another. Furthermore, the theory predicted that for the target absence case, the overall search time should be twice that for the target present case. This required the assumption that subjects searched the display with full memory so that they would not be wasting time with repeated attentional visits to the same item.

Treisman's theory fit the data both qualitatively and quantitatively and in comparison to other areas of psychology, the theory was rich in detailed predictions. For those of us in the field of visual perception and psychophysics (readers of this journal) and who had ambitions to link perception to brain function, FIT was extremely attractive. It meant that simply designed experiments in visual search could reveal new feature maps or at least confirm them. In contrast, topics like object recognition were deemed too complicated and distant for possible neurologizing. Explaining everyday behavior in terms of the known properties of neurons was heady stuff. Furthermore it gave a new possible meaning to the intuitive yet somewhat ill defined concept of attention. Instead of just meaning greater processing, it gave attention a designated computational role, linking features to locations, something that appealed to those of us who wanted something more concrete.

Furthermore, it implied that simple psychological experiments themselves could conceivably play a key role and be on a par with single unit recording itself, since FIT postulated that flat search functions would be diagnostic of a canonical feature map. Julesz's texton theory also posited a similar role for psychological and psychophysical experiments. Thus, psychological experiments alone might find out important facts about the neurological organization of the visual system.

Those of us vision researchers who were drawn to this framework were tempted to think of how feature integration theory might be more specifically realized within particular visual cortical structures with their known receptive field properties. The story fit at least in broad outline, because it was clear that neurons in various parts of the visual cortex were selectively sensitive to features and thus it seemed entirely reasonable that a feature map based on say a given color, a given orientation, a particular direction of motion existed. This would in turn form the substrate for feature maps that could mediate rapid search or popout.

On closer reflection, however, it became clear that some neurons in striate and extrastriate cortex were not tuned to just one feature, but to many and jointly at least to two. For example, striate cortical neurons jointly coded spatial frequency and orientation (De Valois et al., 1982; Campbell et al., 1969). If one thought of other visual cortical areas, there were additional examples of other cells tuned to joint dimensions such that optimal stimulation along each dimension would yield the greatest responses. In MT for example, cells were tuned both to direction of motion and binocular disparity (Maunsell & van Essen, 1983). Thus, it seemed reasonable to extend FIT's popout account to include conjunctions, since unique conjunctions in a given cortical area might be thought themselves to have the same properties as single features in terms of a singular pattern of activation.

With this in mind, Nakayama and Silverman (1986) reasoned that since binocular disparity was coded in many cortical areas jointly with other dimensions such as motion and color, conjunctions of binocular disparity with dimensions might also enjoy such pre-attentive status. As such, they constructed a search array in three dimensions using stereopsis such that for a given rectangular grid of search elements, each element in the array could either be "in back" or "in front" as defined by binocular disparity. If one put all of the blue colored distractors in back and red ones in the front, a conjunctive target could be defined as a red behind (as shown in Fig. 2a) or a blue in front. According to FIT's original formulation, these targets should not popout since no map coding a single feature would be uniquely activated. In contrast, Nakayama and Silverman found that there was no increase in search time. The same flat functions obtained for stereo-motion (SM) and stereo color (SC) conjunctions (Fig. 2b).

Nakayama was very pleased with these results, for although they contradicted a central tenet of feature integration theory in its original form, they could be seen as strongly supportive of FIT if one widened the definition of feature maps to include joint coding of features by neurons. Initially, he saw the result as the obvious and necessary extension of FIT, given the need to reconcile the theory with existing neurophysiological knowledge. It was deemed a happy marriage because feature maps of two dimensions plainly existed and the flat search functions supply ample evidence for such an extension of Feature integration theory. Furthermore, by doing more experiments using conjunctions, one might distinguish those features in neurons that were jointly coded from those which were not.

Already, however, there were a couple of disquieting facts in this report that seemed to go against feature integration theory. First (referring to Fig. 2B), while there was no increase in response time with increasing distractor number, the Y intercepts for the stereo motion (SM) and the stereo color (SC) were very high, approximately 1500–2000 ms. This kind of behavior was never seen in previous "popout" experiments where constant reaction times were more on the order of 500 ms or far less. Later studies would show that popout defined more precisely could be very slow (Bravo & Nakayama, 1992) so this may have been a less serious objection.

More worrisome was a phenomenological aspect of the experiment, namely that it seemed to the subjects that they were attending to each depth plane selectively, perhaps sequentially and that they were then finding the odd color within a given attended depth plane. This observation was odds with the unique spirit of FIT, which postulated that attention was *not* required for popout. So, while objectively it seemed that the results extended the results of feature integration theory to conjunctions, subjectively at least it seemed violate one of its central tenets – the assertion that there was some form of privileged communication between early visual processing and the site where search performance is determined, thus bypassing attention. However, it took more formal experimentation with more objective measures to deal with these issues, one to show that binocular disparity cannot be seen as a defining feature in explaining conjunctive search and another to show that attention was clearly required for simple feature search.

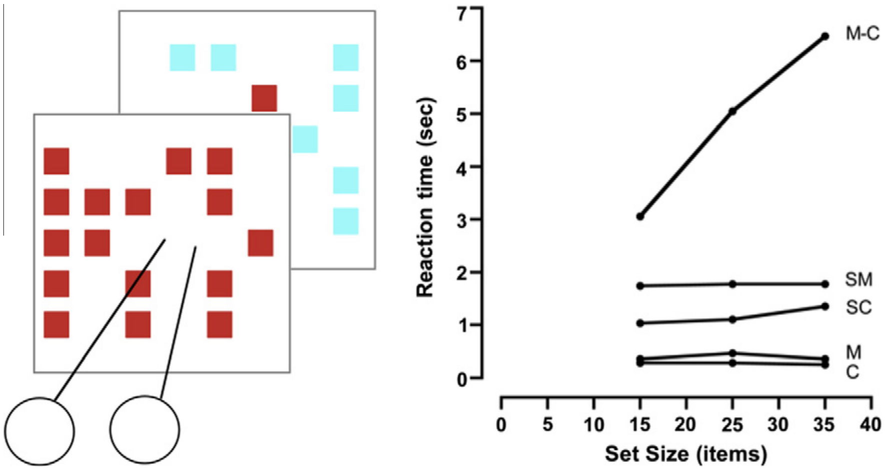

Fig. 2. (A) Schematic description of the Nakayama (1986) conjunctive stereo color (SC) visual search display. The observer's task was to find the odd color for a given depth (red target in back or blue in front). Elements in both the front and back are fully visible (not occluded) and are only seen as opaque planes here for illustrative purposes; (B) Reaction times plotted against number of distractors show constant reaction time flat functions for conjunctions of stereo and color (SC) and stereo and motion (SM). In comparison, another conjunction task, color and motion (CM) showed very steep slopes. Simple tasks, finding a single color alone (C) or motion alone (M) showed much shorter reaction times and flat functions.

Replicating and further characterizing the phenomenological observations accompanying the Nakayama and Silverman visual search experiment, He and Nakayama (1995) had observers attend to elements having the middle–depth values (defined by binocular disparity) where there was an odd colored target amongst distractors and where the same target colors were also present in the nearer and farther distances (Fig. 3a and b). For the case depicted in Fig. 3a, it was possible to attend to the middle disparity, ignoring closer and farther disparities. Reaction times were short and there was little increase in reaction time with increasing distractor number, thus replicating Nakayama (1986). However, binocular disparity differences alone were not sufficient to assure "popout". Note that in this case, the search elements are not only at a particular disparity but the elements are arranged so that each element is coplanar with all other elements at this disparity. Thus, all middle disparity elements form a well defined surface. However, note that in Fig. 3b, the elements in a search plane were not co-planar. In this case, He and Nakayama (1995) reported that it was very difficult to selectively attend to the middle disparity elements, requiring great effort and search times were consistently higher by almost half a second. Therefore binocular disparity was not sufficient for this efficient focusing of attention to a middle plane because there was the additional requirement of co-planarity.

Not only was binocular disparity not *sufficient*, it was not *necessary* as He and Nakayama (1995) also showed that a common binocular disparity was irrelevant for selecting planes by using uniquely colored targets in horizontally oriented planes (compare Fig. 3c and d). These easily attended planes having co-planar elements did not even share a common disparity. Simply put, whether an array could be effectively attended was not determined

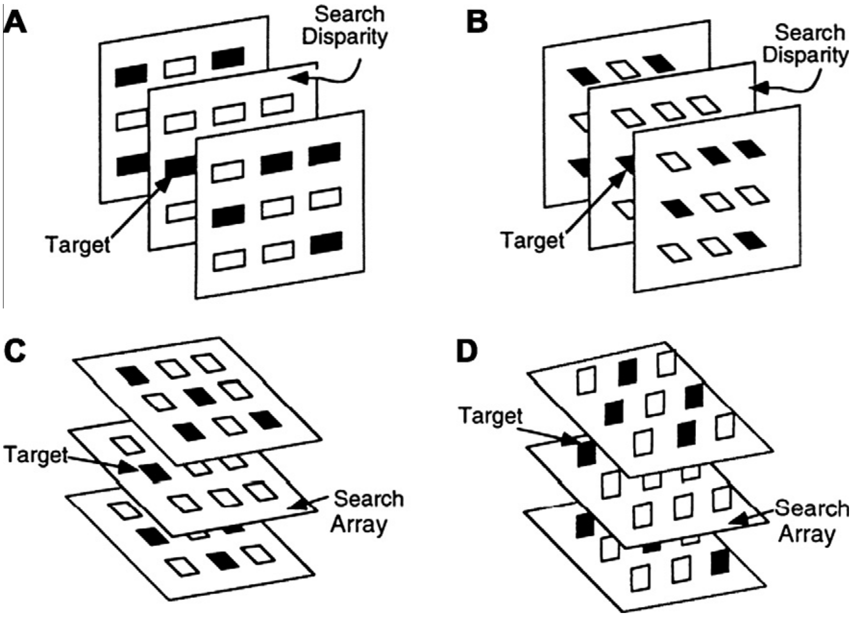

Fig. 3. Visual search task where the subject is to find the odd color target at the intermediate depth value, where the elements comprise a well formed surface (A) and where they do not (B). Fig. 3C shows the stimulus configuration where its easy to search a horizontal surface spanning many values of binocular disparity with co-planar elements whereas in Fig. 3D, it's very hard to do this because the elements do not lie within this horizontal plane. (Reprinted with permission from He and Nakayama (1995)).

by any scheme of joint codings by elementary receptive fields, but by whether they comprised a well formed surface, a higher-level form of representation (see also Nakayama, He, & Shimojo, 1995).

Joseph, Chun, and Nakayama (1997) directly addressed another central tenet of FIT, that attention can be bypassed and can have a direct route to those processes which report target detection in a visual search task. In a dual task attentional blink paradigm subjects were instructed to report a cued letter in an RSVP letter stream while also doing a "popout" task, determining whether an otherwise easily visible, oddly oriented Gabor patch was present. There was a dramatic diminution of performance for several 100 ms in the dual task condition, showing that indeed the attentional resources captured by the RSVP task essentially abolish the phenomenon of popout. This confirmed in a very different way the notion that detecting popout of a singleton in a display required attention, dispelling the notion of an "attention free" privileged route.

3. More doubts about the importance of low-level features

The basic assumption as well as a motivating force for early researchers in the study of visual search was the notion that the simple receptive properties of neurons in retinotopic cortex had a privileged role in perception. There were many good reasons to make this optimistic assumption, as psychophysics had been successful in identifying color channels corresponding to the ones measured physiologically and the most recent identification of spatial frequency channels with characteristics similar to cortical neurons (Blakemore, 1969; De Valois & De Valois, 1988) was similarly encouraging. Thus, there was no denying that the existence and characteristics of early mechanisms could be felt in some carefully designed psychophysical studies. The question remained as to whether these mechanisms are the ones that will reveal themselves more generally, specifically in other tasks such as visual search. Over the period of a few years it became clear that these putative elementary processes played little or no direct role in determining visual search performance.

First there were well known effects of practice. For any given visual search pattern, many studies have shown that accuracy improves and reaction times decline over time. If we assume that visual search were to be determined by many invariant simple features, this is an aberrant finding. Against this, it was argued that in some situations subjects were learning some extraneous aspect of the task, getting used to the procedures, developing a strategy. Such arguments do not apply to the experiments of Wang, Cavanagh, and Green (1994), who relied on extreme differences in familiarity with patterns which should otherwise be equivalent at the level of low-level feature maps. They showed that detecting a target shaped as a among a shape as distractors was very easy, whereas finding among distractors was much harder. Each had essentially the same configuration of elements in terms of an early visual representation and should activate early maps similarly. Because the and resemble the numerals 2 and 5, a highly practiced pattern for westerners, did it now qualify as a feature? Similar findings showing a strong role for familiarity were seen for Chinese and Persian writing symbols in these populations.

Another important and otherwise puzzling finding was the very different search rates for differently shaded cubes (see Fig. 4a and b reported by Enns and Rensink (1990)). Small differences in the ordering of brightness on cube-like patterns have a huge effect on visual search performance. Here the difference is very subtle if one thinks only of image feature combinations but is much more interpretable if these cubes are seen as a three dimensional scene, where in the normal viewing situation a planar surface of cubes is lit from above as shown in Fig. 4A. Here the task was very easy. This should be compared to the case where they are lit form below (Fig. 4B), something occurring very infrequently and thus leading to very slow search. The difference in performance was dramatic as shown by the data presented on the right panel of Fig. 4.

In other experiments, researchers questioned the importance of image features at all. Employing binocular disparity to manipulate depth, He and Nakayama (1992) showed that one could find Ls among reversed L's with ease if the L's were put in front of squares so that they were perceived as L shapes (Fig. 5A). However, if depth was manipulated such that the L image region was seen behind the squares, visual search became extremely difficult. In this case, the L image shape was not seen as an L but as a square, amodally completing behind an occluding square (Fig. 5B). Thus L's and reversed L's are seen as much less distinct by virtue of them both looking more like squares. This means that visual search was not determined by properties of the image, i.e. low level representations, but by higher order ones, the inferred shape of partially occluded objects.

All of these findings taken together strongly indicate that image properties are not determinative of visual search performance, that simple cortical receptive fields are not able to account for the results (see also Nakayama and Joseph, 1998).

All of these results point out that the provisional architecture suggested by FIT or any other image-based feedforward architecture (we dub this Visual Architecture 1.0) is an inappropriate one to encompass all the findings of visual search. Nevertheless, Treisman's FIT was a great service to vision researchers as it helped the research community to embrace and for many to reject the simple models of vision implied by the single unit physiology of the Hubel and Wiesel era. A new conception was needed, but it did not come out as a single framework or theory, but several related ones (Hochstein & Ahissar, 2002, Edelman, 1987; Lamme, 2003). It is still today in embryonic form, but discernible is a core which de-emphasizes canonical detectors, ignores the "binding" problems and allows for very high level processing to occur very early in time. While not all of this emerging framework is needed to contextualize visual search, a brief sketch of some of these developments provides an alternative perspective.

4. Vision Architecture 2.0

There were a number of antecedents that set the stage for the birthpangs of what may eventually become Vision Architecture 2.0.

5. Short latency responses in higher visual areas

Hubel and Wiesel and those that followed broadly assumed a serial model of vision, with many stages in the visual pathway that made sense, both in terms of simple neural connectivity as well as in terms of the properties of neurons at progressively higher stages. This serial hierarchical framework was the implicit operating assumption for years and many models of vision mirrored this assumed structure. However, there were always well known bits of information that raised doubts, reminders that the visual system was unlikely to operate in this way. Most well known for decades was the existence of back projections in the visual system. For example, the number of centrifugal fibers projecting back from the visual cortex to the lateral geniculate nucleus exceeds the number of centripetal fibers by an order of magnitude. This was troubling indeed if one were to think of a strict feed-forward scheme.

Also, there were several related pieces of information which indicated that the flow of information might not precede from the simple to complex. First, were studies of the latencies of responses in various visual areas. Despite the hierarchy suggested

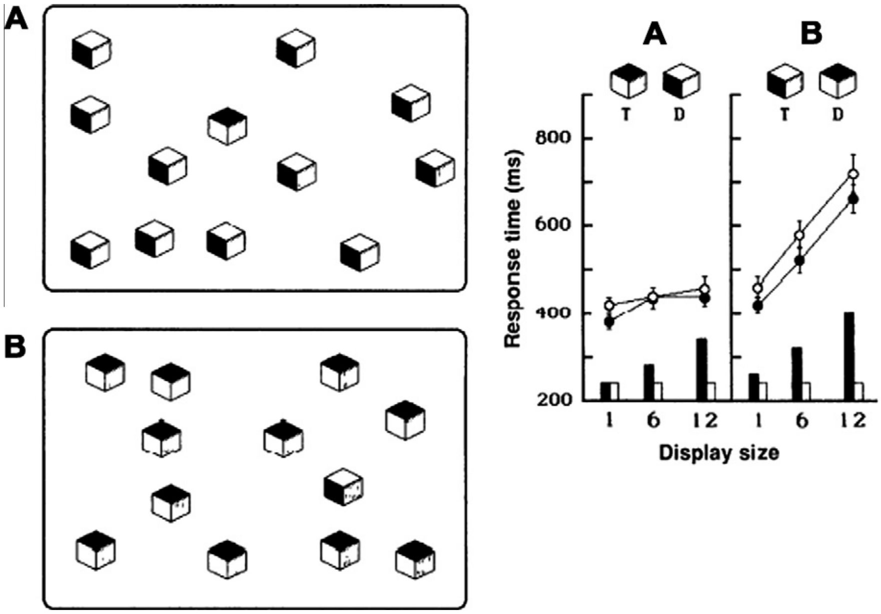

Fig. 4. Comparison of visual search arrays and performance for odd shaded cubes, showing examples which mimic top lit cubes (A) and where they do not (B) Search for odd target in (A) is relatively easy with little increase in reaction time as increasing display sizes in comparison to odd target in (B) which is much harder to find. (from Enns & Rensink, 1990).

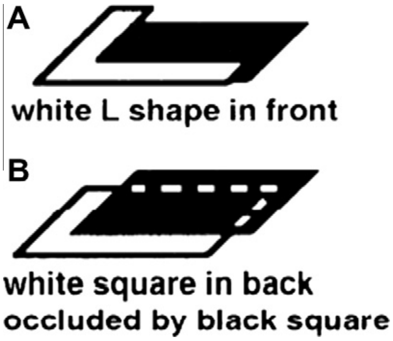

Fig. 5. Comparison of the perception evoked by an L shaped image region presented stereoscopically in front (A) or in back (B). When present in front, it looks like an L, when in back, it is perceived as a square, amodally completing behind the black occluder. Correspondingly, in a visual search task finding the single oppositely oriented "L" in B is much slower than in A, because both L and reversed Ls look like squares (He & Nakayama, 1992).

by van Essen and colleagues, many reports showed that the latencies of cortical neurons did not correspond to the positions on hierarchical maps (Schmolesky et al., 1998). Response latencies in putatively much later areas (frontal eye fields, MT) could respond more quickly than many cell types in earlier cortical structures. This was not surprising given the well known existence of parallel pathways (magnocellular and parvocellular) having different response times which indicated that there should be no simple time sequencing of signals as one moved up the visual hierarchy, that cells at the highest levels might respond equally quickly as cells in low level retinotopic areas.

5.1. Rapid high level discrimination

Consistent with these findings and further challenging a simple serial model, single units in higher brain areas show surprisingly short latency responses to abstract visual categories. For example, there exists a differential response in neurons in cortex which respond with latencies of 100 ms or less to food vs non-food (Fukuda, Ono, Nishino, & Sasaki, 1986). Simon Thorpe and colleagues have also found comparable evidence for this in humans using behavioral and ERP studies. Most dramatic is the recent work (Kirchner & Thorpe, 2006) where using a two choice eye movement task, complex shapes can be discriminated extremely quickly, with the correct choice occurring in 130 ms. Similar work using faces shows correct latencies of even shorter duration, less than 110 ms (Crouzet, Kirchner, & Thorpe, 2010). Furthermore, even very high level discrimination tasks such the discrimination of emotional scenes can be accomplished in less than 100 ms (Maljkovic & Martini, 2005a). This means that contrary to the assumptions of a simple serial model of visual processing, there is much temporal mixing of signals in the visual system, with fast and slow streams co-existing.

5.2. Unconscious vision

The study of unconscious vision has had a long and irregular career, gaining acceptance only after many decades. It started in the 1950s under the moniker of the "New Look in Perception" (Bruner & Goodman, 1947). Inspired by concepts of psychoanalysis then prevalent, McGinnies (1949) reported that observers failed to perceive taboo words (usually with sexual content) as they had special mechanisms that protected them from gaining access to conscious awareness (perceptual defense). These claims were challenged, however, by those who pointed out the likelihood of response bias and criterion changes (see Goldiamond, 1958).

Traditional vision researchers paid little attention to these claims as they measured thresholds – that minimum amount of stimulus strength to elicit "conscious vision". Researchers made the pragmatic assumption that signals had to pass through a series of stages and with appropriate stimulus manipulations there would be some critical, limiting stage (retinal or cortical) whose characteristics would be revealed (Blakemore, 1969; Westheimer, 1965). Conscious visual experience was just a handy indicator. They were not interested in looking for examples where the choke point (or bottleneck), between seeing and not seeing would be determined elsewhere, mediated possibly by attention and perhaps at the level of object recognition. This would have to wait until it became more than evident that unconscious visual processing was something that could not be ignored.

This occurred quietly over the past three or four decades, as the vision research community became slowly aware of widespread examples of visual processing *without* awareness. It started with the phenomenon of blind sight, showing correct behavioral responses in cortically blind fields (Weiskrantz, Warrington, Sanders, & Marshall, 1974). This was followed by the phenomenon of adaptive motor behavior to unseen changes in target position (Bridgeman, Lewis, Heit, & Nagle, 1979). This was followed by the even more surprising phenomenon of masked priming, the fact that words were recognized faster if preceded by unseen versions of them or related words (Marcel, 1983). Despite the range and scope of these studies and their important implications, such results were either ignored or regarded with sceptism.

A decade later, however, the floodgates opened, with reports of a diverse set of phenomena showing very strong effects of unseen visual stimuli, including effects under binocular rivalry, RSVP streams (the attentional blink), continuous flash suppression, fusion suppression, backward masking and object substitution masking. In all of these cases, there were noticeable behavioral consequences, but the stimulus which triggered these effects remained invisible. Although there were many skeptics at first, the sheer volume of unconscious effects grew too large to ignore. The traditional methods of threshold psychophysics was obviously not the only approach to studying the performance characteristics of the visual system. Visual threshold psychophysics could delineate the "seen" from the "unseen", but this only heightened ones curiosity as to a new zone of mystery, the extent and characteristics of the "unseen".

The territory of the "unseen" turns out to be surprisingly large as further characterized by behavioral and human neurophysiological approaches (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006). A briefly flashed word rendered "unconscious" by masking, nevertheless accelerates the recognition of the same word vs other visible words presented subsequently. In addition, the magnitude of this effect occurs independent of case, indicating the activation of a more abstract representation of words not directly tied to visual appearance. In addition, these unseen words activated areas of the fusiform gyrus of the temporal lobe as measured by fMRI (Dehaene et al., 2001). Unseen faces rendered unconscious either by continuous flash suppression or other forms of binocular rivalry showed strong fMRI activation in FFA and amygdala (Williams, Morris, McGlone, Abbott, & Mattingley, 2004).

All of these experiments suggest that high-level meaningful aspects of a stimulus go very far into the visual system. In fact they can go all the way to the motor output. Well known are Milner and Goodale's results (1995) showing that subject DF can correctly grasp objects that she cannot report the shape of. This suggests a direct linkage of vision processing to motor output, mediated by a "dorsal" visual system hypothesized to determine behavior without the need for consciousness. But the level of unconscious processing that can directly lead to motor responses may not only be confined to the dorsal visual pathway. For example, flashed and masked unseen words, which are presumably processed in the ventral pathway, can nevertheless elicit quick responses. This was demonstrated in a study where subjects were to identify the color of a verbally defined item (spinach, blood) by touching a red or green square on the screen. Here an unseen incongruent word "red" or "green", presented just before, initiated incorrect trajectories towards the colored button corresponding to the unseen word, rather than the color of the seen item (Finkbeiner, Song, Nakayama, & Caramazza, 2008). This study shows that unseen symbols (in the form of otherwise arbitrary letter shapes) having no inherently visual relation to colors, can by linguistic association drive behavior. Equally dramatic is the fact that unseen erotic stimuli masked in a binocular rivalry paradigm can direct attention to and away from different parts of the visual field depending on sexual orientation (Jiang, Costello, et al., 2006).

There are additional aspects to the emergence of Vision Architecture 2.0, such as the issue of re-entrant processing and its relation to consciousness that go beyond the scope of the present topic.

6. Implications for visual search

It appears from these general findings about vision that the system very quickly processes information at a deep level of meaning and that much of this can be automatic and unconscious. These results are at odds with the Vision 1.0 idea that there are elemental primitives of vision, features or textons, that later and presumably slower stages (needing attention) can assemble.

The existence of significant depth of processing for both unconscious vision and rapid vision is also mirrored in visual search findings. It's not the shape of the visible fragment of an L itself, but the inferred square shape of the L-shaped region amodally completed behind (Fig. 5). Additionally its not just collection of cubes but a surface array of cubes lit from above (Fig. 4).

Perhaps it is no coincidence that neither rapid high-level visual recognition nor efficient visual search are convincingly understood in terms of primitive canonical textons or features, that they both require relatively high-level visual representations. We argue that both are problems in the recognition of objects and patterns. Each requires that we distinguish one thing from another, one object class from another in the case of object recognition and whether a target is present or absent, essentially one pattern vs another, in a search task.

As such, we briefly survey work on object recognition to see how it might shed light on visual search. An important issue in object recognition has been the problem of invariance. How is it that we can recognize an object or groups of objects under so many poses, lighting, exemplars? Sixty years ago there was much talk of cardinal object recognition units (Konorski, 1967) or grandmother cells (Gross, 2002) units that fire for that object under a wide range of circumstances. While not repudiating the notion of cardinal units, recent emphasis has been on understanding how one class of objects or one specific object can be discriminated from others.

Implicitly retaining the notion of possible decision units for recognition, DiCarlo and Cox (2007) considered object recognition as a problem of linear classification. Thus a neural response to a given sampled image of an object can be considered as a vector (in a very high dimensional space) and that all appearances of this object (under different guises, poses, lighting, etc.) trace out a manifold of such vectors in this space (see Fig. 6a and b). For object recognition, the task is to find a way to separate one manifold from another. Assuming a linear classification scheme, one must find that set of discriminative hyper-planes that best separates the two (as shown in Fig. 6a). This sidesteps the problem of invariance, allowing each object to have a range of neural representations yet also allowing it to be distinctive (DiCarlo & Cox, 2007). The advantage of this conception lies in its compatibility with the longstanding "integrate and fire" models of neurons, that they can sum up excitatory and inhibitory influences (from many dimensions) and come up with an output, and that in doing so they can serve as object memory units.

However, as DiCarlo and Cox point out, for most neural representations of a stimulus, there is no possible hyper-plane cut to separate the two objects, as their corresponding representational manifolds are hopelessly entangled (as shown in Fig. 6b). The challenge for the visual system, as outlined in this framework, is to have some well coded and appropriate representation (as depicted

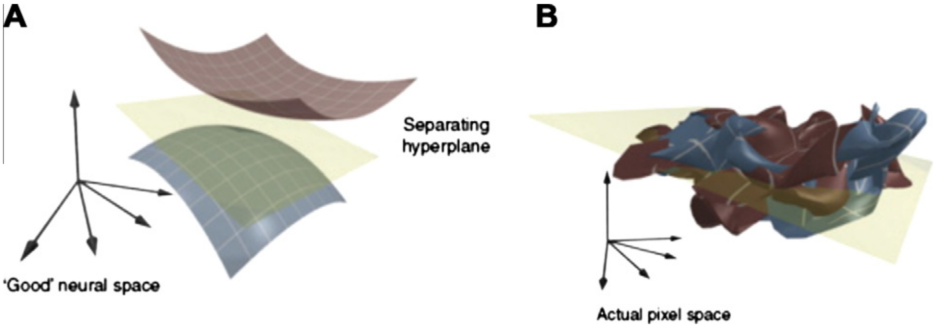

Fig. 6. Manifolds corresponding to untangled (a) and tangled (b) representations of two objects. Manifolds represent multiple image samples from a given object or class of object, occupied by changes of view, lighting, different exemplars, etc. Note that for the untangled representation (A), where object recognition is possible, there exists a linear cut (a hyper-plane) that separates the two objects over these widely varying conditions. For the tangled representation (B), no such plane exists and object recognition is not possible. (reproduced with permission from DiCarlo and Cox (2007).

in Fig. 6a) so that such a cut is possible. This issue of recoding the image has been a longstanding topic in vision (Barlow, 1961). With respect to object recognition, some treatments suggest recoding of the image by concatenating elementary units inspired by Hubel and Wiesel era neurophysiology (Fukushima, 1980; Riesenhuber & Poggio, 1999) whereas other approaches have attempted to develop image fragment detectors that are particularly effective in best discriminating the identity of real world objects under various views (Ullman, Vidal-Naquet, & Sali, 2002). The latter approach has appeal in that the units look more like the kinds of receptive fields found in IT cortex (Tanaka, 1996), although its often hard to make this determination unequivocally. Whatever the mechanism, however, its clear that IT neurons have many if not all of the characteristics to do the job of general object recognition.

Work by Hung, Krienen, Poggio, and DiCarlo (2005) suggests that it's likely that this is accomplished by progressively better coding of stimuli at successive visual stages. By combining the aggregate multiple recordings over time from single units in monkey IT cortex, Hung et al. (2005) simulated a cortex of 256 sites and asked whether their outputs (using machine learning methods) could be used to correctly classify eight different object classes. The results were dramatic, a classification score of 94% was obtained. This means that well placed hyper-planes in these linear spaces were extremely successful in assigning stimuli to the various categories, thus accomplishing the job of object classification. Similar results were obtained for specific objects themselves. In contrast, by taking the outputs of earlier stages than IT (V1 neurons for example) and doing the same procedures, little performance above chance was obtained. This means that in the information flow from V1 to IT cortical areas, there is a radical re-representation of data so that the classification of objects becomes possible. Furthermore the process is accurate even with very short latencies (100 ms) and with minimal duration of the response (12.5 ms). This synthetic exercise thus argues for the DiCarlo and Cox (1997) concept of untangling in understanding object recognition.

Can we now apply this untangling concept to the domain of visual search? Twenty years ago, D'Zmura (1991) came up with an essentially identical formulation in a visual search experiment where he had subjects look for a designated target color among a set of two colored distractors. Using a variety of targets and distractors arrayed in two dimensional hue space, he concluded that the key determinate as to whether the search would be easy (no increase in RT with increasing distractor number) was the ability to draw a straight line in the color space separating targets and distractors. Thus, in Fig. 7a the target can be easily dissociated from the distractors whereas this is not the case for Fig. 7b. Further work by Bauer, Jolicoeur, and Cowan (1996) confirmed D'Zmura's results and made the stronger case that it was unlikely to be target distractor distance in the space, but the ease with which the targets and distractors could be cleaved by a single line. Thus we have a foreshadowing in miniature (with just two dimensions) of the higher dimensional untangling hypothesis of DiCarlo and Cox (2007). The line suffices to do the job of linear classification of the hyper-plane.

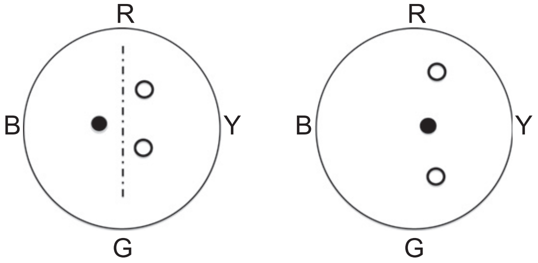

Fig. 7. Color space (YB, RG axes) framework to understand why visual search is easy sometimes (left) but hard at other times (right) when a single color target (solid symbol) is searched among two distractor colors (open symbols). It's easy when a straight line can be between the distractors and targets as on the left (redrawn from D'Zmura, 1991).

These two contributions, DiCarlo and Cox (2007) for object recognition and D'Zmura (1991) for visual search, highlight their similarities as well as their differences. Both rest on the same idea of linear classification to do pattern recognition. For object recognition, it's for one object at a time with a high number of dimensions required for the classification. For visual search, it's for many objects at a time, with very few distinguishing dimensions (or detail). Thus each represents the extremes of plausible trade-offs between dimensions vs objects (or details vs scope). We suggest that there are finite computational resources for any pattern recognition task for a given system, one can have lots of dimensions or lots of objects but not both. In Fig. 8, we illustrate this limitation by showing the conceivably possible relations between scope (objects) on the abscissa and dimensions (detail) on the ordinate. Assuming a logarithmic scale, the negative 45° line traces out the limits of a system with fixed capacity (since all points along this line have the same product). For any pattern classification that can be done in single glance, we suggest that points must lie below this 45° line, patterns represented by points above this line exceed the capacity of the recognition system.

Object recognition and visual search thus occupy extreme positions within this feasibility triangle, with object recognition nestled into the upper left corner and easy visual search, in the lower right hand corner. Object recognition is usually described for one object at a time and thus allows for recognition using the maximum numbers of dimensions. This is the DiCarlo and Cox formulation where manifolds must be untangled for object

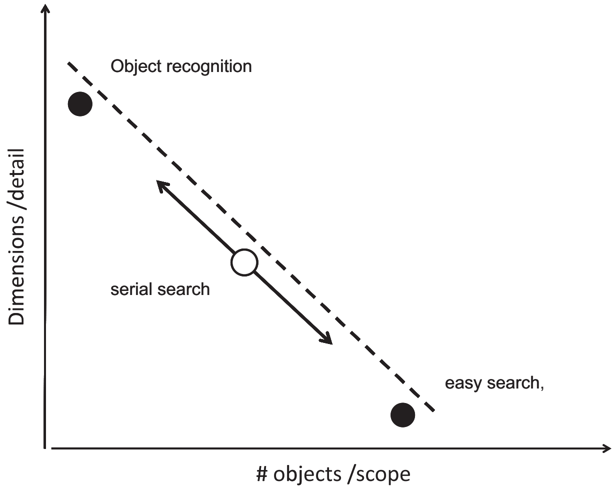

Fig. 8. Computational resource framework to understand the relation between object recognition and visual search. The negative dashed 45° diagonal line in this log–log plot depicts a line of fixed computational resources (the product of number of objects and dimensions). The area below this line represents a feasibility triangle where pattern recognition can occur in a single glance. Object recognition is customarily based on single objects at a time with a maximum number of dimensions is in the upper left corner, whereas easy visual search spans many items with few dimensions and resides in the lower right corner. More difficult visual search can occupy intermediate positions (serial search) depending on the number of recognition dimensions required.

recognition to proceed. Easy visual search on the other hand, as described by Treisman and also by D'Zmura, is formally the same, a pattern recognition task but over many objects and larger spatial extents with only few distinctive dimensions required, dimensions which represent efficient codes and could be complex (Ullman et al., 2002) It thus correspondingly lies in the lower right corner. Search that cannot be done in one glance or in a single attentional fixation (Nakayama, 1990) must undergo the same pattern recognition process repeatedly but can only do this with a subset of the whole display. The size of the subset is dictated by the number of dimensions required for this identification (Alvarez & Cavanagh, 2004). More difficult searches require more dimensions and thus a reduced spatial extent of pattern recognition and points are correspondingly shifted leftward on this diagram. Less difficult searches requiring fewer dimensions, can operate over more objects and larger areas and are shifted rightward. As such, the locus of serial search tasks in our framework traces out diagonal line within the feasibility triangle (also shown in Fig. 8).

The trade-off between scope (number of dimensions) and scale (number of objects) allows one to situate any visual search task within this framework – meaning that there is an essentially infinite combination of intermediate cases where tasks partake characteristics of object recognition and/or visual search to varying degrees. Somewhat elastic is the concept of an object or an item and this can be easily influenced by efficient coding. For example, the Enns and Rensink easy search for a plane of top lit cubes (Fig. 5) might be considered as a single object or surface rather than numerous cubes. Thus search is easy, not because it just had a low numbers of dimensions, but because the display itself might be considered as a single object and thus placed further to the left on the feasibility triangle than other easy search tasks. We also suggest that examples of rapid scene categorization described earlier (Kirchner & Thorpe, 2006) can be situated within this same framework, sharing the locus with easy search in the lower right corner in Fig. 8.

Summing up. We have been developing the view that visual search has had a disproportionate role in the history of vision science, that it was certainly popular and continues to be so. Our view is that some of this enthusiasm has been misplaced, based on the often unexamined assumption that visual search is closely tied to elementary canonical detectors so seemingly well established for early visual areas. However, by conceptualizing visual search in the category of object recognition tasks, we hope to stimulate research along these lines which could better characterize visual search as well as object recognition itself.

7. Prediction and memory in visual search

So far we have stressed the link between visual search and emerging ideas about visual function, tracing the history of visual search, showing that it mirrors ideas about how we understand the visual system.

Thinking farther afield and into the future, it seems that visual search has and will continue to have a prominent role in understanding a broader range of issues, ones that have long been a concern to researchers who have considered visual search in a more applied framework – from searching for lost men and submarines at sea (Koopman, 1980), for predators looking for things to eat (Poulton, 1890) and more recently, for security screeners looking for weapons in baggage and doctors looking for rare tumors (Wolfe, Horowitz, & Kenner, 2005).

A wide range of endeavors have considered visual search in this context and we trace this back to some of its origins. We mentioned Bernard Koopman earlier, a then well known mathematician who developed the notion that search for targets in a featureless field (searching for a ship in a wide ocean) was likely to be akin to a random process, because the uncertainty due to imperfect visibility and other factors limited the probability of detecting the target in a single sample (Koopman, 1956b). As such, search is closely related to the classic sampling with replacement from an urn with easily derivable predictions. The probability of success as a function of samples (or time) would be exponential. Koopman's predictions were confirmed in great detail in well controlled visual search experiments in Philadelphia's Franklin planetarium, where small targets were to be found against a featureless sky under varying durations and spatial extent (Krendel & Wodinsky, 1960). The results were entirely consistent with a constrained random walk with no memory. These dramatic findings indicated that very simple ideas, describable mathematically, could make exact predictions about complex behavior.

However, most of the visual search displays concocted by psychologists were more structured, they comprised target items and distractors and it would seem reasonable that memory would be important. In fact, perfect memory was the underlying assumption for Treisman and Gelade's prediction of the 2:1 ratio of the slope of reaction time functions for serial search, where target present trials would have half the slope as target absent ones.

However, many factors are involved in deciding when to terminate a serial search (Chun & Wolfe, 1996) and so the role of memory cannot be so easily isolated and assessed. For some search tasks, with moderate numbers of items, search may appear to be random, as demonstrated by Horowitz and Wolfe in a very clever study: by replacing items continuously under intermittent stimulation, they showed no costs, thus no evidence of memory for serial search (see also Kristjansson, 2000). However, a simple tendency to avoid resampling from a just visited location, known as the inhibition of return phenomenon (Klein, 1988; Klein & MacInnes, 1999), can mimic very closely the performance of an ideal searcher (Najemnik & Geisler, 2005), suggesting that what is necessary and sufficient for optimal behavior might be a form of memory of an elementary kind, just keeping track of what happened one trial back.

Situating visual search in its ecological context of foraging may provide further clues. Ecologists have long known that foraging animals do not behave as random searchers, rather the time series of their feeding choices depart from randomness in very distinctive ways. Birds feeding on grains of different types tend to produce long streaks of pecks to the same grain type (Fig. 9), unlikely to arise from random sampling (Dawkins, 1971). Predators feeding on cryptic prey tend to eat disproportionately more the most common type of prey ((Bond, 2007; Bond & Kamil, 2002), the phenomenon is called frequency-dependent predation. In laboratory situations, where different targets are rewarded with uneven frequencies, animals tend to match the frequency of reward by choosing targets in proportion to the frequency of rewards received according to Herrnstein's (1961) matching law.

In human visual search behavior, such as security screening for baggage control at airports or inspection of organ tissues in the medical profession, rarely encountered items such as weapons or tumors are missed much more often than expected by chance (Wolfe & Van Wert, 2010; Wolfe et al., 2005). As such, the statistical frequency of stimuli must play a strategic role in visual search.

What mechanism is responsible for these frequency dependent behaviors? A crucial observation is that statistical frequency is inextricably linked to the degree of repetition of the stimulus: stimuli that occur more frequently also repeat more often (Maljkovic & Martini, 2005b). In the early days of the Cognitive Revolution during the 1960s it was realized that repetition was the crucial determinant of the dependence of choice reaction time on stimulus frequency, formally known as Hick's law (Kornblum, 1968, 1969; Schweickert, 1993). The further realization that an elementary form of memory needs to be invoked to explain sensitivity to repetition arose from two distinct, but arguably linked fields of study.

First were the findings of Maljkovic and Nakayama (1994, 1996), demonstrating that the reaction time for finding a target singleton in a three-items, pop-out search display depends on the history of the stimulus sequence. By adopting a reverse correlation analysis familiar to system neuroscientists, they computed memory kernels for the target's features (Fig. 10A). These kernels indicated that each encounter with a target speeds up the future 8–10 responses to targets with the same features and at the same time slows down responses to targets with different features.

Secondly, a similar history effect was independently discovered in the study of responses to rewards. Hunter and Davison (1985) studied the dynamics of responses to changes in reinforcer ratios in concurrent variable-interval schedules and computed kernels demonstrating that pigeons' response ratios following an abrupt change in reinforcer ratio reached steady state in about five sessions. The more recent studies by Newsome and colleagues (Corrado, Sugrue, Seung, & Newsome, 2005) and Glimcher (Lau & Glimcher, 2005) used techniques identical to those of Maljkovic and Nakayama to compute kernels for a single reward, demonstrating that in the monkey a single rewarding event influences several future choices according to a kernel function that is suspiciously similar to that found in pop-out search (Martini, 2010). As such, these studies demonstrated that reward frequency is computed locally through a form of leaky integration of past reward encounters (Fig. 10B).

Both phenomena, i.e. sequential dependencies in pop-out search and in reinforcement schedules, can be described as forms of implicit short-term memory (Maljkovic & Nakayama, 2000), yet such description may appear to have limited explanatory power, being essentially a restatement of the phenomenon. What is needed is a functional explanation.

Progress in this direction appears more advanced in the field of reward processing than in visual search, where ideas from animal behavior, physiology and machine learning have congealed into the coherent explanatory framework of reinforcement learning (Sutton & Barto, 1998). Particularly important in this context has been the concept of temporal difference learning, where choices are driven by value functions that are updated on every trial by adding a weighted prediction error, the difference between the actual reward and the previously computed value. Theories of this kind have recently received much attention following the discovery that dopaminergic activity may be the physiological substrate of the prediction error (Schultz, Dayan, & Montague, 1997).

In contrast, memory kernels in visual search have been discussed mostly within the confines of attention guidance mechanisms. The idea that exposure to a particular feature increases the attentional weight to that feature in future choices has been suggested in diverse species performing diverse tasks: pigeons pecking at grains of different colors (Dawkins, 1971), blue jays feeding on polymorphic moths (Bond & Kamil, 2002) and humans searching for a pop-out target (Maljkovic & Nakayama, 2000), to name a few examples. The question remains as to why attention should be weighted more heavily towards features

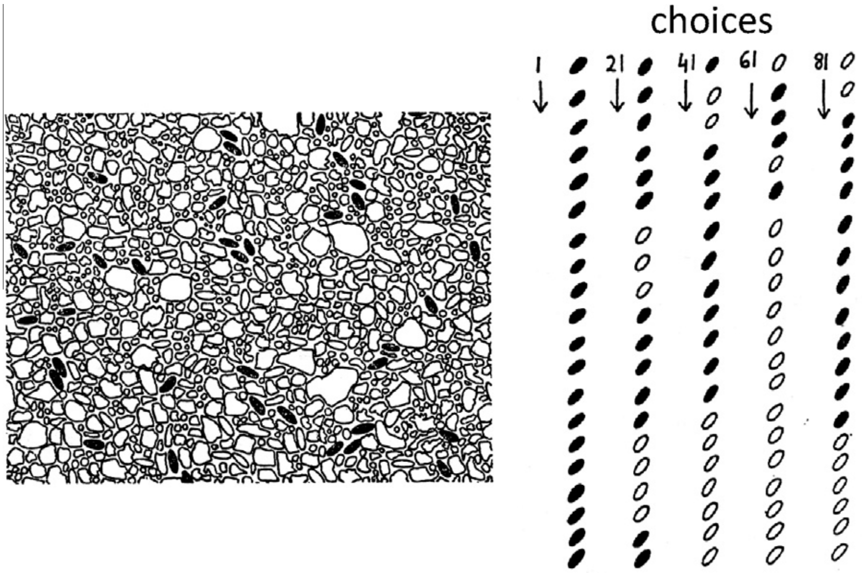

Fig. 9. Right panel shows long sequences of pecks at light or dark grains during the feeding of a chick, where randomly placed grains are on the ground (as shown on left). Data obtained from video tape (from Dawkins, 1971).

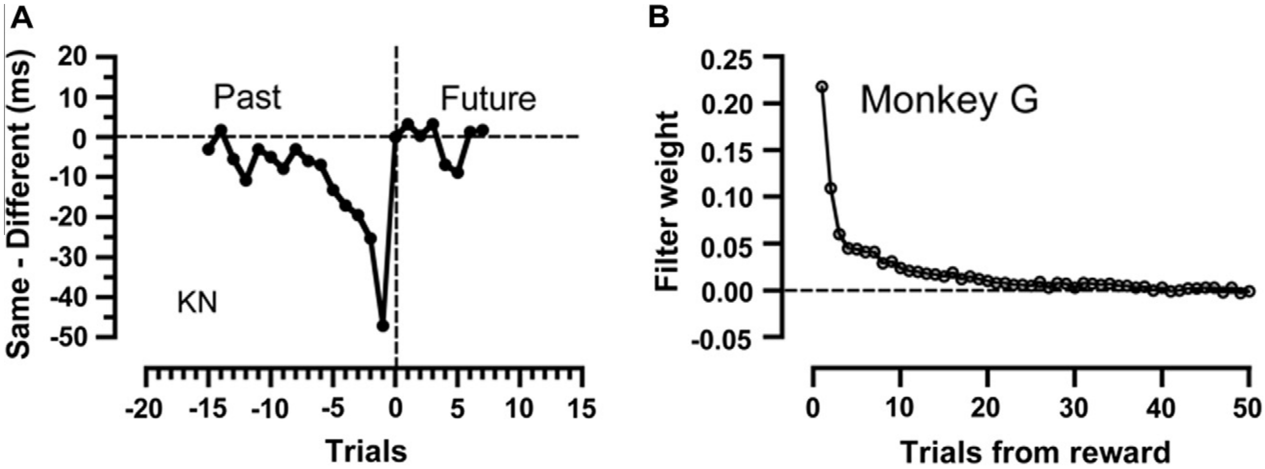

Fig. 10. Behavioral memory kernels derived from studies on the deployment of attention and from reinforcement schedules. In (A), we see one of the first examples using reverse correlation in behavioral studies to obtain the effect of a particular color target on reaction times to the same vs different colored targets across trials (Maljkovic & Nakayama, 1994). The facilitation (negative reaction time difference) for the same color target is plotted as a function of trials since presentation. Note that there is a long lasting effect of a single presentation, such that a target presented as many as 15 trials in the past still has a measurable influence on the current trial. As a cross check on the method, its clear that there is no effect of future trials on the present. In (B), we see a similar dependence of reward in a two choice reinforcement schedule, where reward in the current trial can have effects long into the future (redrawn from Corrado et al., 2005). In both cases, the decay functions have similar forms.

or targets that have just occurred: why isn't attention diminished by repetition?

One possible answer has to do with prediction. Consider the goal of finding a target among distractors: is there anything that can be learned from the history of trials that will improve the chance of finding the target quickly? This amounts to learning regularities or departures from randomness that can lead to anticipation of the next trial, such as uneven target probabilities or temporal correlations in the targets' sequence. The priors for assuming that prediction should be possible must be very high, because completely random phenomena are very uncommon in nature. Most natural time series are significantly correlated, as demonstrated by ubiquitous 1/*f*-type spectra (Hurst, Black, & Simaika, 1965), and resources tend to be spatially clustered within homogeneous patches (Taylor, Woiwod, & Perry, 1978). A searcher with an inbuilt tendency toward prediction will then try to exploit these regularities by default.

In temporal difference learning discounting past and predicting future rewards are two aspects of the same learning algorithm. We suggest that a similar learning process operates also in visual search: prediction is attained by weighting attention to one feature in the upcoming trial proportionally to its previous encounters. As such, fluctuations in attentional weights related to feature repetitions have much in common with fluctuations in motivational salience induced by rewards (Maunsell, 2004).

8. Searching into the future

Long-term predictions are hazardous, especially in evolving systems such as scientific research. Nevertheless, with over two thousand articles on visual search, one thousand of these in the past decade, it's unlikely that researchers will abandon visual search arrays in studying vision and behavior. If our personal sampling and reflection of this literature has captured significant underlying themes, we suggest ways that further studies of visual search could reap benefits.

As visual search is more widely understood as a species of pattern recognition along with other object recognition problems, conceptual advances in each field should be mutually beneficial, progress in object recognition should inform our understanding of visual search and vice versa.

Similarly, as visual search captures choice behavior in a microcosm, with its easily characterizable properties it may offer opportunities for understanding reinforcement and value, and hopefully it may benefit from the large research effort currently undergoing in decision making.

Acknowledgment

This research was supported in part by RCUK Fellowship to P. Martini and NIH Grant (R01 EY01362) to K. Nakayama.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*(2), 106–111.

Barlow, H. B. (1953). Summation and inhibition in the frog's retina. *Journal of Physiology, 119*(1), 69–88.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith (Ed.), *Sensory communication*. Cambridge, MA: MIT Press.

Bauer, B., Jolicoeur, P., & Cowan, W. B. (1996). Visual search for colour targets that are or are not linearly separable from distractors. *Vision Research, 36*, 1439–1465.

Blakemore, C., & Campbell, F. W. (1969). On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology, 203*, 237–260.

Bond, A. B. (2007). The evolution of color polymorphism: Crypticity searching images, and apostatic selection. *Annual Review of Ecology Evolution and Systematics, 38*, 489–514.

Bond, A. B., & Kamil, A. C. (2002). Visual predators select for crypticity and polymorphism in virtual prey. *Nature, 415*(6872), 609–613.

Bravo, M., & Nakayama, K. (1992). The role of attention in different visual search tasks. *Perception & Psychophysics, 51*, 465–472.

Bridgeman, B., Lewis, S., Heit, G., & Nagle, M. (1979). Relation between cognitive and motor-oriented systems of visual position perception. *Journal of Experimental Psychology-Human Perception and Performance, 5*, 692–700.

Bruner, J. S., & Goodman, C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology, 42*, 33–44.

Campbell, F. W., Cooper, G. F., & Enroth-Cugell, C. (1969). The spatial selectivity of the visual cells of the cat. *Journal of Physiology, 203*, 223–235.

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.

Corrado, G. S., Sugrue, L. P., Seung, H. S., & Newsome, W. T. (2005). Linear-nonlinear-Poisson models of primate choice dynamics. *Journal of the Experimental Analysis of Behavior, 84*(3), 581–617.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision, 10*, 1–17.

Dawkins, M. (1971). Shifts of "attention" in chicks during feeding. *Animal Behaviour, 19*(3), 575–582.

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends Cogition Science, 10*(5), 204–211.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Natural Neurosciencence, 4*(7), 752–758.

De Valois, Russell L., Albrecht, Duane G., & Thorell, Lisa G. (1982). Spatial frequency selectivity of cellsin macaque visual cortex. *Vision Research, 22*(5), 545–559.

De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. New York: Oxford University Press.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences, 11*, 333–341.

D'Zmura, M. (1991). Color in visual search. *Vision Research, 31*, 951–966.

Edelman, G. M. (1987). *Neural Darwinism:* The theory of neuronal group selection. New York: Basic Books.

Enns, J. T., & Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science, 247*, 721–723.

Finkbeiner, M., Song, J. H., Nakayama, K., & Caramazza, A. (2008). Engaging the motor system with masked orthographic primes: A kinematic analysis. *Visual Cognition, 16*, 11–22.

Fukuda, M., Ono, T., Nishino, H., & Sasaki, K. (1986). Visual responses related to food discrimination in monkey lateral hypothalamus during operant feeding behavior. *Brain Research, 374*(2), 249–259.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics, 36*, 193–202.

Goldiamond, I. (1958). Indicators of perception. 1. Subliminal perception, subception, unconscious perception – An analysis in terms of psychophysical indicator methodology. *Psychological Bulletin, 55*, 373–411.

Gross, C. G. (2002). The genealogy of the "Grandmother Cell". *The Neuroscientist, 8*, 512–518.

He, Z. J., & Nakayama, K. (1992). Surfaces vs. features in visual search. *Nature, 359*, 231–233.

He, Z. J., & Nakayama, K. (1995). Visual attention to surfaces in 3-D space. *Proceedings of the National Academy of Sciences, 92*, 11155–11159.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior, 4*, 267–272.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*, 791–804.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology, 148*, 574–591.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology London*, 215–243.

Hung, C. P., Krienen, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310*, 863–866.

Hunter, I., & Davison, M. (1985). Determination of a behavioral transfer function: White-noise analysis of session-to-session response-ratio dynamics on concurrent VI schedules. *Journal of the Experimental Analysis of Behavior, 43*(1), 43–59.

Hurst, H. E., Black, R. P., & Simaika, Y. M. (1965). *Long-term storage, an experimental study* (pp. xiv, 145 p.). London: Constable.

Jiang, Y., Costello, P., et al. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences of the United States of America, 103*(45), 17048–17052.

Joseph, J. S., Chun, M. M., & Nakayama, K. (1997). Attentional requirements in a "preattentive" feature search task. *Nature, 387*, 805–808.

Julesz, B. (1964). Binocular depth perception without familiarity cues. *Science, 145*, 356–362.

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature, 290*, 91–97.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46*, 1762–1776.

Klein, R. (1988). Inhibitory tagging system facilitates visual search. *Nature, 334*(6181), 430–431.

Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science, 10*(4), 346–352.

Konorski, J. (1967). *Integrative activity of the brain*. U. Chicago Press.

Koopman, B. O. (1956a). The theory of search. I. Kinematic bases. *Operations Research, 4*(3), 324–346. CR – Copyright &#169; 1956 INFORMS.

Koopman, B. O. (1956b). The theory of search. II. Target detection. *Operations Research, 4*(5), 503–531.

Koopman, B. O. (1957). The theory of search. III. The optimum distribution of searching effort. *Operations Research, 5*(5), 613–626.

Koopman, B. O. (1980). *Search and screening: General principles with historical applications* (pp. x, 369 p.). Elmsford, N.Y.: Pergamon Press.

Kornblum, S. (1968). Serial-choice reaction time: Inadequacies of the information hypothesis. *Science, 159*(3813), 432–434.

Kornblum, S. (1969). Sequential determinants of information processing in serial and discrete choice reaction time. *Psychological Review, 76*(2), 113–131.

Krendel, E. S., & Wodinsky, J. (1960). Search in an unstructured visual field. *Journal of the Optical Society of America, 50*, 562–568.

Kristjansson, A. (2000). In search of remembrance: Evidence for memory in visual search. *Psychological Science, 11*(4), 328–332.

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7*, 12–18.

Lau, B., & Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *Journal of the Experimental Analysis of Behavior, 84*(3), 555–579.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the IRE, 47*(11).

Maljkovic, V., & Martini, P. (2005a). Short-term memory for scenes with affective content. *Journal of Vision, 5*(3), 215–229.

Maljkovic, V., & Martini, P. (2005b). Implicit short-term memory and event frequency effects in visual search. *Vision Research, 45*(21), 2831–2846.

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition, 22*(6), 657–672.

Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. The role of position. *Perception & Psychophysics, 58*(7), 977–991.

Maljkovic, V., & Nakayama, K. (2000). Priming of pop-out: III. A short-term implicit memory system beneficial for rapid target selection. *Visual Cognition, 7*(5), 571–595.

Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15*, 197–237.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Freeman.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London Series B-Biological Sciences, 207*, 187–217.

McGinnies, E. (1949). Emotionality and perceptual defense. *Psychological Review, 56*, 244–251.

Martini, P. (2010). System identification in Priming of popout. *Vision Research, 50*, 2110–2115.

Maunsell, J. H. (2004). Neuronal representations of cognitive state: Reward or attention? *Trends in Cognitive Sciences, 8*(6), 261–265.

Maunsell, J. H. R., & van Essen, D. C. (1983). Functional properties of neurons in middle temporal area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology, 49*, 1148–1167.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434*(7031), 387–391.

Nakayama, K. (1990). The iconic bottleneck and the tenuous link between early visual processing and perception. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 411–422). Cambridge University Press.

Nakayama, K., & Joseph, J. S. (1998). Attention, pattern recognition and popout in visual search. In R. Parasuraman (Ed.), *The attentive brain* (pp. 279–298). Cambridge: MIT Press.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature, 320*, 264–265.

Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher level vision. In S. M. Kosslyn & D. N. Osherson (Eds.), *Vision in invitation to cognitive science* (pp. 1–70). M.I.T. Press.

Poulton, E. B. (1890). *The colours of animals: Their meaning and use especially considered in the case of insects*. London: Kegan Paul. 360 pp.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019–1025.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593–1599.

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., et al. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology, 79*, 3272–3278.

Schweickert, R. (1993). Information, time, and the structure of mental events: A twenty-five-year review. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance 14: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 535–566). Cambridge, MA: The MIT Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (pp. xviii, 322 p.). Cambridge, Mass: MIT Press.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19*, 109–139.

Taylor, L. R., Woiwod, I. P., & Perry, J. N. (1978). Density-dependence of spatial behavior and rarity of randomness. *Journal of Animal Ecology, 47*(2), 383–406.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5*, 682–687.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception & Psychophysics, 56*, 495–500.

Weiskrantz, L., Warrington, E. K., Sanders, M. D., & Marshall, J. (1974). Visual capacity in the hemianopic field following a restricted occipital ablation. *Brain, 97*(4), 709–728.

Westheimer, G. (1965). Spatial interaction in human retina during scotopic vision. *Journal of Physiology, 181*, 881–894.

Williams, M. A., Morris, A. P., McGlone, F., Abbott, D. F., & Mattingley, J. B. (2004). Amygdala responses to fearful and happy facial expressions under conditions of binocular suppression. *Journal of Neuroscience, 24*(12), 2898–2904.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature, 435*(7041), 439–440.

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology, 20*(2), 121–124.