The level of attention: Mediating between the stimulus and perception

Jeremy M Wolfe

Center for Ophthalmic Research

Brigham and Women's Hospital and Harvard Medical School

221 Longwood Ave, Boston, MA 02115

wolfe@search.bwh.harvard.edu

Current conceptions of visual processing make good use of the metaphor of levels of vision. At the very least, there are meaningful distinctions to be made between early vision, mid-level vision, and high-level vision. Without getting too committed to the details, early vision is a level of local processing of simple stimulus attributes like the orientation and motion of line segments. Perhaps these can be considered to be the atoms of vision. If so, then mid-level vision is concerned with the molecules - larger pieces put together out of the early vision atoms. Like the "wetness" of water, these mid-level molecules may have properties that are not easy to predict from their early vision precursors. Examples might include Gestalt observations about the whole being greater than the sum of its parts (Kellman, 1998; Rock & Palmer, 1990) or the work of Adelson (refXX - this volume), Gilchrist (refXX - this volume) and others on the apparent brightness of surfaces. At a still higher level, the molecules of mid-level vision give rise to recognizable objects.

Once upon a time, perhaps in the first flush of excitement about single-unit recordings in the visual system (Barlow, 1995), we might have thought of these levels in a fairly straight-forward, hierarchical manner. The atoms (lines) made the molecules (corners, junctions) that made the object (grandmother)(Lettvin, 1995). Even then, it was a bit difficult to believe in this effectively unidirectional story given evidence for massive feedback from apparently higher levels in the visual nervous system to earlier stages (reviewed in Lamme, Super, & Spekresijse, 1998). Moreover, even if everything was feeding forward, it became clear that not everything was being allowed to pass unimpeded to the higher levels. You could not simultaneously recognize all of the objects in the visual field. Somewhere there was a gate or a filter or bottleneck that was permitting some information to flow to higher levels while other information was blocked, perhaps lost. Attention altered the filter or moved the bottleneck around the visual field. Reading provides a clear example. Even if letters of text are made big enough to be read without eye movements, reading proceeds in a serial fashion as one word after another is somehow selected and processed. The stimulus on the retina might remain constant but the contents of the later stages had to be changed by an act of selecting some stimuli and ignoring others. Neisser (1967) distinguished between levels of processing that were preattentive in which all input could be processed in parallel and attentive levels that processed only a selection of the available input at any one time.

The purpose of this paper is to revisit the role of attention in mediating between levels of visual processing. The heart of the argument will be that attention governs the surprisingly narrow gate between visual perception and the visual stimuli that give rise to that perception. To make that argument, this paper will review several lines of research, drawn largely from the visual search literature. This evidence will be seen to back us into a corner. The data can be used to make a good case for the argument that we only "see" one object at a time (cf (Mack & Rock, 1998a, 1998b; O'Regan, 1992; Rensink, 2000; Simons & Levin, 1997). How can we reconcile this evidence with our subjective impression of a rich visual world, populated with many objects? We will assume that the there is a rich physical world *out there* to be seen. We will argue, against some current thinking, (e.g. O'Regan & Noe, 2001), that perceivers experience a rich perceptual representation of that world. The curious, perhaps counterintuitive argument of this chapter is that the representation is in rather limited contact with the stimuli that give rise to that representation. Thus, the cat you "see" may not be quite the same as the cat in the world. Indeed, that physical cat might be gone from the physical scene. It is through the narrow gate of attention that the perceived cat is linked to the cat in the world. This is a version of the venerable thought that we infer the visual world (Brainard, Wandell, & Chichilnisky, 1993; Freeman, 1994; Helmholtz, 1924; Nakayama & Shimojo, 1992). This paper will examine the role of attention in the maintenance of that inference.

1) Evidence for the parallel processing of visual features

Visual search experiments provide one line of evidence for the parallel processing of some visual features. In a typical visual search experiment, subjects look for a target among a variable number of distractors. A useful dependent measure is the reaction time (RT), the time to respond that, 'yes', a target is present or 'no', it is not. The slope of the function relating RT to the number of items (set size) is a measure of the efficiency of a search task. For a limited set of stimulus attributes, that slope is near zero when the target is defined by the presence of a feature. There are, perhaps, a dozen dimensions that will support this sort of highly efficient search (reviewed in Wolfe, 1998). Figure One shows some examples: 1a - luminance polarity (Enns & Kingstone, 1995; Gilchrist, W., Jane, & Heiko, 1997; O'Connell & Treisman, 1990) which may or may not be the same dimension as color (Carter, 1982; D'Zmura, 1991; Green & Anderson, 1956; Treisman & Gormican, 1988); 1b - orientation (Foster & Ward, 1991; Moraglia, 1989; Nothdurft, 1992); 1c - size (Aks & Enns, 1996; Cavanagh, Arguin, & Treisman, 1990; Williams, 1966), ld - line termination (Treisman & Gormican, 1988) or, conversely, closure (Elder & Zucker, 1993), 1e - curvature (Wolfe, Yee, & Friedman-Hill, 1992) (see also Kristjansson & Tse, 2001), and 1f - various aspects of 3D structure [Enns, 1990 #2059; Enns, 1990 #2321; Enns, 1992 #3191; Rensink, 1998 #5017; Kleffner, 1992 #3201; Sun,

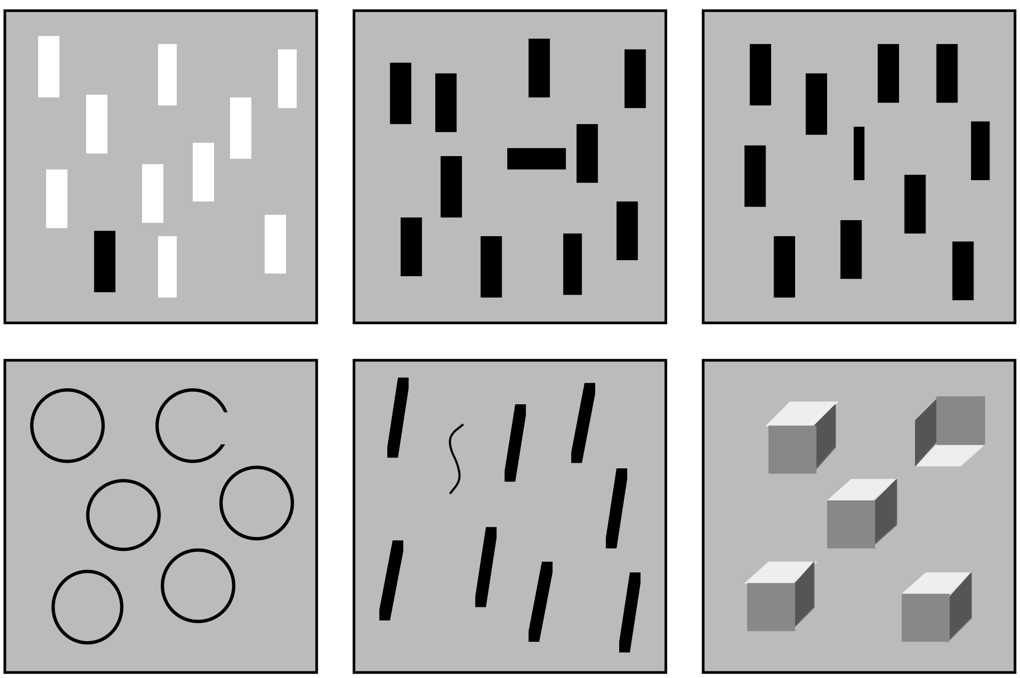

Figure One: Examples of simple searches for "basic features".

1996 #4190].

It is not entirely trivial to determine what is and is not a basic feature in visual search. For example, it is sometimes possible to search for conjunctions of two or more features very efficiently (Theeuwes & Kooi, 1994; Wolfe, 1992). It is unlikely that this means that the visual system possesses a mechanism for parallel processing of conjunctions of orientation and luminance polarity, for instance. More likely, the system can guide attention simultaneously toward items with the target orientation and the target polarity (Wolfe, Cave, & Franzel, 1989). The claim that any particular attribute is a basic feature is strengthened when there is converging evidence, notably from visual search asymmetries (Treisman & Gormican, 1988; Treisman & Souther, 1985; Wolfe, 2001) and texture segmentation (Beck, 1966; Julesz & Bergen, 1983; Wolfe, 1992). For present purposes, the important point is that there is a limited set of features that appear to be processed in parallel, across the visual field and that appear to be available to guide the deployment of attention.

Basic features and early vision

It is important not to confuse "basic features", as defined within the visual search literature, with early vision features as assessed with classical psychophysical and electrophysiological methods. There are many important features that make it clear that basic features in visual search are not properties of cells in early stages of visual cortical processing (e.g. V1).

1) The lists of features are different.

There are a number of candidates for basic feature status in visual search that are not generally found on lists of early vision features. These might include lighting direction and shading (Braun, 1993; Enns & Rensink, 1990a; Kleffner, Polichar, & Ramachandran, 1990 ; Ramachandran, 1988; Rensink & Cavanagh, 1993; Sun & Perona, 1996a, 1996b), binocular lustre (shininess) (Wolfe & Franzel, 1988) and a variety of depth cues(Enns & Rensink, 1990a, 1990b; Previc & Naegele, 2001; Rensink & Cavanagh, 1994; Sun & Perona, 1996a).

2) Preattentive basic features can be created as "second order" stimuli.

For example, it is easy to find a vertical target among horizontal distractors. These vertical and horizontal stimuli can be created by simple luminance differences between stimulus and background but they can also be based on texture or grouping of other elements (even other oriented elements Bravo & Blake, 1990). The oriented regions can be defined by attributes such as motion, color, or stereopsis (Cavanagh et al., 1990).

3) Coding of preattentive basic features appears to be quite coarse.

One can measure "just noticeable differences" (jnd) with standard psychophysical methods in order to determine when two stimuli can be discriminated at some threshold level. One can measure a different sort of jnd in visual search by measuring the slope of the RT x set size function for a range of differences between target and distractors. A somewhat arbitrary slope threshold can define a preattentive jnd just as a somewhat arbitrary discrimination threshold defines the classic JND. When such experiments have been done in the color (Nagy & Sanchez, 1990) and orientation (Foster & Westland, 1992, 1998), we find that preattentive jnds are much larger than classical jnds.

The effects of distractor heterogeneity reveal another sense in which preattentive basic features are coarsely coded. It is generally true that increasing distractor heterogeneity

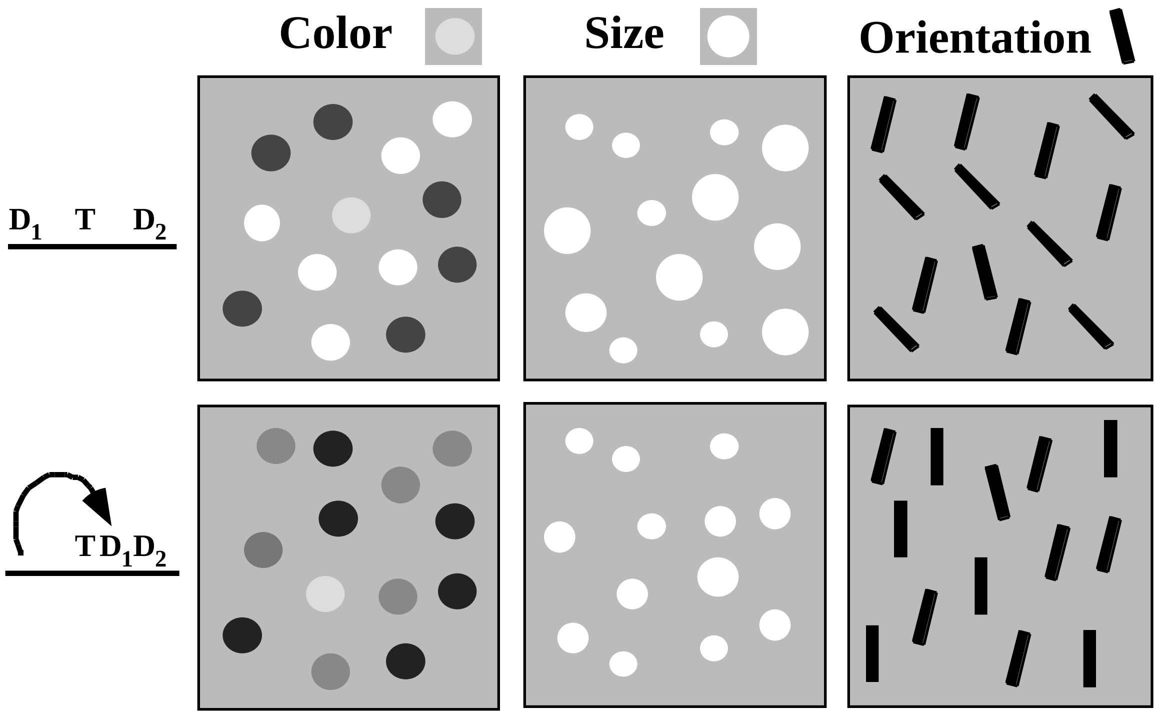

Figure Two: It relatively hard to find targets that are flanked by distractors in feature space (e.g. medium size target among big and small distractors). It is easier to find targets that to one side of the distractors in feature space.

decreases search efficiency (Duncan & Humphreys, 1989). More specifically, search is particularly difficult when the distractors flank the targets in feature space. This is demonstrated in Figure 2.

In the first row of the figure, the distractors flank the target in feature space (e.g. in the size panel, the target is intermediate in size between the large and small distractors.) In the second row, one of the distractor types is changed so that it lies between the target and the other distractor in the feature space. This makes the distractors, on average, more similar to the target but the task becomes easier. This may be clearest in the orientation example, where it is easier to find the line tilted 15 deg left (-15) among 0 and 15 deg distractors (bottom row) than among flanking -45 and 15 deg distractors (top row). Extensive experimental support for this claim has been obtained for color (Bauer, Jolicoeur, & Cowan, 1996; Bauer, Jolicoeur, & Cowan, 1998; Bauer, Jolicœur, & Cowan, 1996; D'Zmura, 1991) and orientation (J M Wolfe & S R Friedman-Hill, 1992; J. M. Wolfe & S. R. Friedman-Hill, 1992; Wolfe, Friedman-Hill, Stewart, & O'Connell, 1992).

4) Coding of preattentive features may be categorical.

In Figure Three, the target orientation on the left is -10 deg and the distractors are +30 & -70. If we add 20 deg to all orientations, we get 10 deg among +50 and -50 deg. In this latter case, the target is *categorically unique*. It is the only "steep" item. The stimuli on the right yield search that is markedly more efficient that the stimuli on the left (Wolfe, Friedman-Hill et al., 1992). In orientation, the preattentive categories seem to be "steep", "shallow", "left" and "right". In size, the categories are probably merely "big" and "small"; in depth, "near" and "far" and so forth.

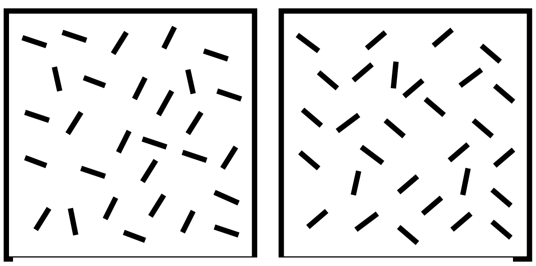

Figure Three: The targets, tilted 10 deg off vertical, are easier to find on the right where they are the only "steep" items than on the left, where they are not categorically unique.

Reverse Hierarchy

Given this set of experimental findings, it is quite clear that preattentive vision is a relatively late abstraction of the visual input. Its sensitivity to second-order stimuli like texture boundaries and its coarse, categorical nature point to a locus beyond the detailed local processing of V1. An interesting problem is posed by this fact. A lot of information seems to have been lost on the way from early vision to the preattentive representation of the visual information that supports efficient visual search. When attention is directed to a stimulus, that information can be recovered. How is that done? In their Reverse Heirarchy Theory (Ahissar & Hochstein, 1997), Ahissar and Hochstein revive a thought that had been discussed earlier, notably in the physiological literature and in some computational models (notably Tsotsos, 1988 , 1993; Tsotsos et al., 1995). Perhaps the role of selective attention is to allow the perceiver to reach back into earlier stages of visual processing and to recover for specific items the details that had been lost in general. In this view, visual attention mediates between levels of visual processing (c.f. re-entrant processes: Di Lollo, Enns, & Rensink, 2000).

The nature of preattentive objects

To understand the implications of this view for an understanding of visual perception, a few more facts are needed. First, let us consider the nature of the preattentive representation in some more detail. One more piece of evidence that preattentive vision comes after early vision is that, in the preattentive representation, the visual scene has been parsed into some sort of objects. In contrast, early vision seems to be concerned with local features and rather minimally with whether features are part of the same or neighboring objects. The primary body of evidence supporting the idea that the preattentive representation contains objects is the evidence that attention tends to select objects (reviewed in Goldsmith, 1998; Tipper & Weaver, 1998). Attention to one part of an object seems to "flow" to other parts of objects (Baylis & Driver, 1993; Egly, Driver, & Rafal, 1994; Tipper, Weaver, Jerreat, & Burak, 1994) (for a physiological analog see Roelfsema, Lammer, & Spekreijse, 1998). Some aspects of object structure seem to be available preattentively (e.g. occlusion Rensink & Enns, 1995). New objects capture attention while equivalent changes in low level features do not (Yantis & Hillstrom, 1994; Yantis & Jonides, 1996).

A precise definition of preattentive "object" has eluded the field to date. In part because it seems clear that the same entity (e.g. "nose") can be an object at one moment and a part of an object at another. Further complicating matters, items that might each be considered to be an object may be grouped to form a different objects for the purpose of visual search (Bravo & Blake, 1990) (see also the 'anthill' phenomenon of Nelson, 1974).

While the nature of an object is not entirely clear, it does seem clear that basic features are rather loosely attached to preattentive objects. Treisman (1982) originally proposed that, prior to the arrival of attention, features were in some sense "free floating" and able to migrate quite widely in a scene, forming "illusory conjunctions" with other features (Treisman & Schmidt, 1982). The notion of completely free floating features seems to overstate the case (Cohen & Ivry, 1989). Treisman (personal communication) argues that the rapid action of attention would provide some coarse location coding for features (see also Cohen & Ivry, 1991). Another hypothesis, tied to the idea that attention is directed to some sort of preattentive objects, holds that features like the color, size, and orientation of an object are loosely "bundled" with the object prior to the arrival of attention (Wolfe

& Bennett, 1997). If an object possesses multiple examples of a single type of feature on a single object (e.g. two or more colors or orientations), the relationship of those features

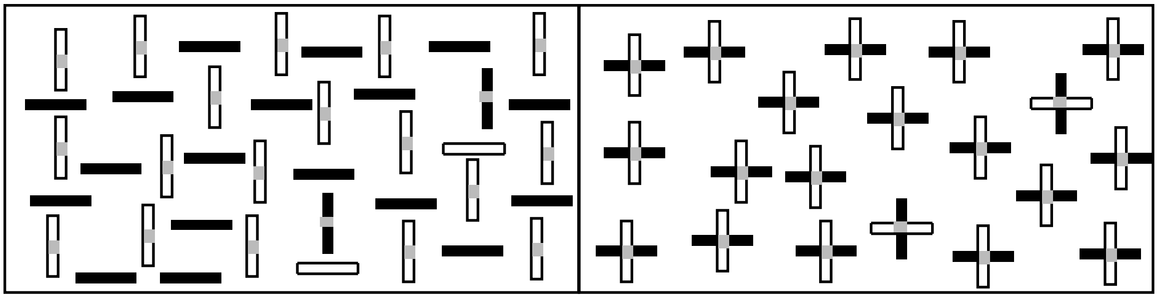

Figure Four: In the left and right panels, search for the black vertical lines (2 in each panel. It is much easier on the left because preattentive feature information can be used to guide attention to objects that a just black and vertical. On the right, all objects have the attributes "black" and "vertical". It requires attention to determine where those features are bound the same part of the object.

to the object and to each other would not be made explicit until attention permitted the accurate "binding" of those features (Treisman, 1996). As an example, consider Figure Four.

On the left, it is quite easy to find black vertical targets because, even if the features are not bound to each other preattentively, attention can be guided to the preattentive bundle that includes the features "black" and "vertical". A version of this conjunction search experiment, using red and green rather than black and white stimuli, yield a target present slope of 5.9 msec/item. In contrast, it is harder to find the black verticals in the right-hand panel of Figure Four (there are two of them). The vertical and horizontal items have been combined into "plusses". The resulting preattentive objects *all* have the features "black, white, vertical and horizontal". These objects only differ when correctly bound. As a result, search is much less efficient: 47.2 msec/item in a red-green version (Wolfe & Bennett, 1997).

To summarize, visual search data suggest that, prior to the arrival of attention, the visual system codes about a dozen basic features in parallel and represents these as bundles of features, loosely aggregated into preattentive objects. Attention to one of these bundles allows the features to be properly bound, making explicit their relationship to one another. This explicit relationship, in turn, makes object recognition possible. There is some evidence for what could be considered "implicit binding" (e.g. Houck & Hoffman, 1986) and even for implicit recognition (e.g. Tipper & Weaver, 1998) but the ability is fairly limited (e.g. Neumann & DeSchepper, 1992). It seems reasonably clear that attention to specific objects in a scene is needed to recognize those specific objects.

Post-attentive vision

What happens after attention has been used to select, bind, and recognize an object? Does its status in visual search change? This can be called the problem of "post-attentive vision" (Wolfe, Klempen, & Dahlen, 2000). In order to address this question, we have performed an extensive series of "repeated search" experiments. These are somewhat different from standard search tasks. In a standard search task, the subject knows that she is looking for a specific target. On each trial, a new set of items appears. It may or may not contain the target. In a repeated search task, the situation is reversed. For a block of trials (several hundred trials in some experiments), all the items in the display remain static, in fixed locations. On each trial, the subject is asked about the presence or absence of a randomly chosen item. Half the time it is an item from the display. Half the time it is absent from the display. The subject simply responds "present" or "absent" in the usual manner.

We have used many different types of stimuli for this task: Letters, novel objects, realistic objects - sometimes embedded in naturalistic scenes. For the example presented here, the stimuli were photorealistic objects provided by Michael Tarr <http://www.cog.brown.edu/~tarr>. A sample display is shown in Figure Five. The actual

Figure Five: Sample stimuli from a *Repeated Search* experiments using realistic objects and auditory probes. The stimuli would remain static across a block of hundreds of trials. On each trial, subjects would hear the name of a target to be searched for (e.g. "bee" - present, "apple" - absent).

items were colored. Subjects were taught the specific names used for each object and all objects were tested for recognizability at all eccentricities used when subjects were fixating a central spot.

In prior studies, we had presented a visual probe or a word on the screen to inform subjects of the target identity for each trial. In this experiment, we used an auditory probe. This allowed the visual stimulus to remain utterly unchanged for a block of trials. Subjects looked at a display like the one shown in Figure Five and heard, for example, "anchor", to which they would have responded in the affirmative or "cow" - negative.

Ten subjects were tested for two blocks of 50 practice and 100 experimental trials at each of four set sizes (4, 6, 10, & 20) Average RTs are shown in Figure Six.

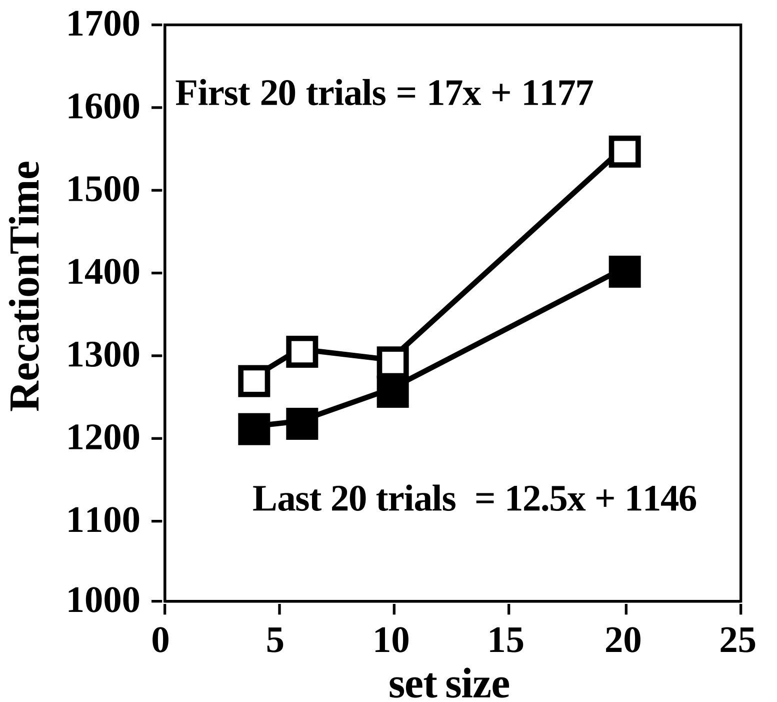

Figure Six: Results of a Repeated Search experiment showing little (statistically insignificant) change in search efficiency after100 trials.

These results are typical. There is little or no improvement over, in this case, 150 trials. The differences seen here are not significant. In many other experiments, we have found little or no improvement in the efficiency of search, as measured by the slope of the RT x set size function. There is no increase in efficiency over the first few trials nor over blocks of up to 350 trials (Wolfe et al., 2000). Apparently, even when only a few items are clearly visible in the field, some bottleneck prevents simultaneous access to all items. You believe that you can see N items. You have memorized N items (at least, if N is relatively small.). Yet, if asked to confirm that you can see one of those items, you behavior depends on the number of items in the display just as it would if you were searching a completely new display.

If the visual stimuli are removed, search actually does become more efficient. The task becomes a memory search task. It is known that memory search can become "automatic" - independent of set size - with several hundred trials of practice (Logan, 1992). In the visual case, however, search does not become automatic. Prolonged exposure to the same, unvarying stimulus does not remove the capacity limitation on search.

One object at a time?

How should we understand the failure of repeated search to become efficient? It may be useful to think about the requirements for object recognition more generally. In order for a visual stimulus to be recognized, its features must be bound (as illustrated by the "plus" example shown in Figure Four). That bound representation must then be linked to an identifying representation in memory. Ian Howard is not recognized as Ian Howard until you have bound the Howard features and linked that bound representation to the Howard representation in your memory. Without that link, the bound representation is *something*

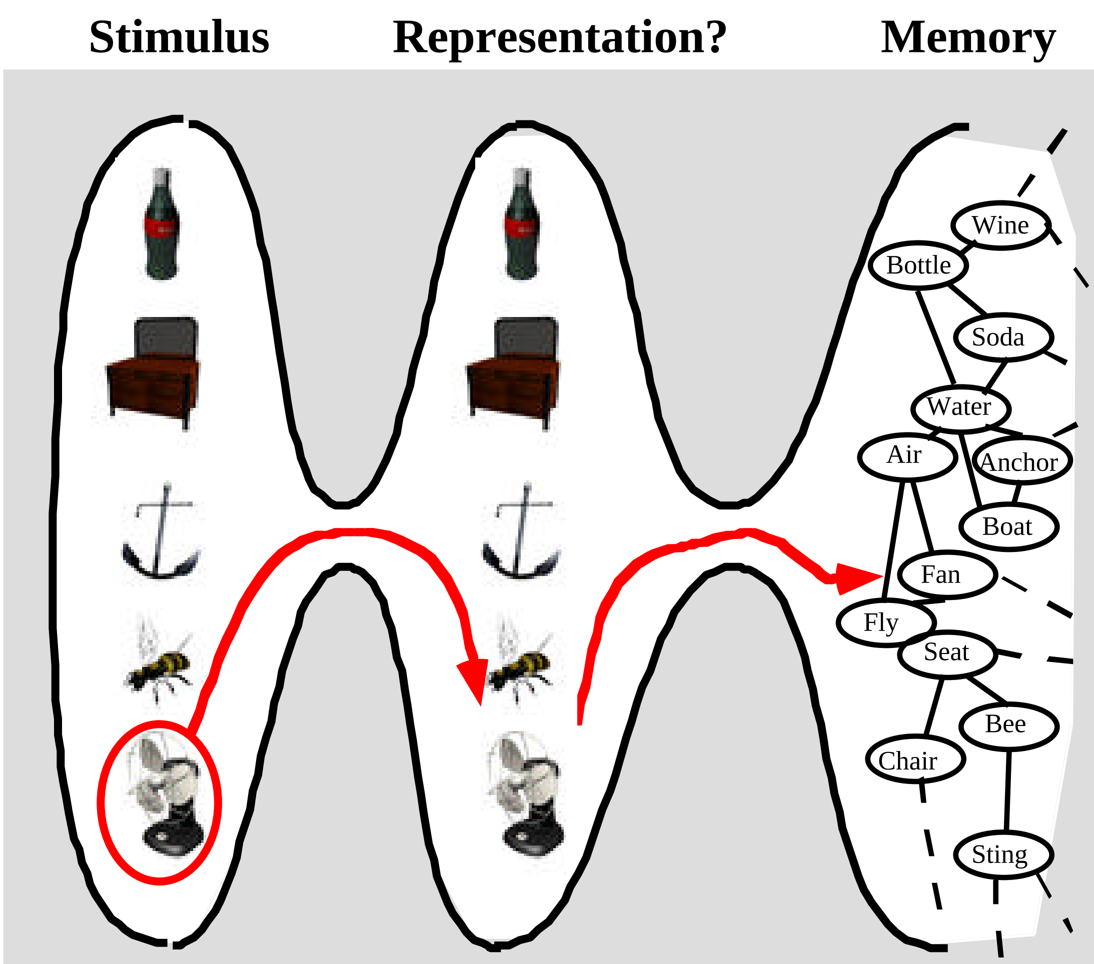

Figure Seven: A cartoon illustrating three broad "levels" in perception. First, the stimulus; second, a hypothetical internal visual representation; and third, memory. Bottlenecks, governed by attention, limit contact between levels.

but it not recognizably Ian. The *repeated search* experiments suggest that either the binding or the linking operations - or both - are limited to a single object at a time. This is shown in cartoon form in Figure Seven. Like most cartoons, this one is not intended to be taken too seriously or too literally. It simply illustrates the idea that there is a stimulus containing a vast amount of information. We each possess a memory containing a vast amount of information. In addition, we *seem* to have a representation of the stimulus that is, itself, rich with information. The reality of that internal representation is discussed in the next section. For the present, the central point is that contact between levels is very restricted. (Note that even though the "stimulus" and "representation" are identical here, this should not be taken as a claim that we create a faithful representation of the outside world.)

The cartoon shows the fan as the attended item. If the bee was replaced by a button while the fan was attended, the "change blindness" literature tells us that an observer would not notice until attention happened to be directed to the object that was previously the bee (Rensink, 2000; Simons & Levin, 1997). In a change blindness experiment, observers see one versions of a scene change into another and report on the nature of the change. If something is done to hide low-level transients that would cue the location of a change, observers prove to be very bad at reporting the change. Objects can appear and disappear and, yet, the change can go unreported until attention happens to be directed to the relevant object. Apparently, attentional bottlenecks keep observers from noticing changes that occur, literally, right before their eyes.

Returning to the cartoon, one can speculate that these bottlenecks serve to prevent the confusion that might arise from crosstalk if there were multiple links between levels. Without a bottleneck, an observer might not be sure which object was being recognized as a fan and which as a bee. Not that, though the cartoon shows bottlenecks between stimulus and representation and between representation and memory, current data are not adequate to determine the number of bottlenecks - only to assert that that must be bottlenecks.

What do we actually see?

The argument for not much.

Data from a variety of different paradigms have been used to argue that the internal representation of the visual world is, at best, very sparse. We cannot faithfully integrate information across saccades (Blackmore, Brelstaff, Nelson, & Troscianko, 1995; Grimes, 1996; Henderson, 1997; Irwin, 1996; Irwin, Yantis, & Jonides, 1983; Irwin, Zacks, & Brown, 1990) though some memory can guide eye movements (e.g. Carlson-Radvansky, 1999; Hollingworth & Henderson, 2001; Karn & Hayhoe, 2000). Even without eye movements, the change blindness literature shows that we are very poor at detecting changes in images if the transients produced by those changes are hidden (e.g. Rensink, O'Regan, & Clark, 1996; Rensink, 2000; Simons & Levin, 1997; Simons, 2000). During on-line tasks, we seem to continually go back to acquire information (Ballard, Hayhoe, & Pelz, 1995; Ballard, Hayhoe, & Pook, 1995; Hayhoe, Bensinger, & Ballard, 1998). We fail to report very basic properties of the visual scene if they are unexpected and unattended (Mack & Rock, 1998a, 1998b). Finally, as noted above, the repeated search data suggest that we do not have simultaneous access to multiple bound, linked, and recognized objects.

Given this body of information, it has been argued that we only see currently attended items (Mack & Rock, 1998a, 1998b) or that we preserve only the minimal information needed for a just-in-time visual system (Ballard, Hayhoe, & Pelz, 1995; Rensink, 2000) or that the world, itself, is the representation with no need for an internal representation (O'Regan & Noe, 2001).

The argument for a rich representation

The difficulty is that people think that they see *something* and, if one wants to understand visual perception , it is ultimately unsatisfying to say that the experience does not exist. So, what do we see? On the one hand, we do not see the stimulus in any very direct form. Even if we did not concern ourselves with attention, phenomena like binocular rivalry (Blake, 1989; Breese, 1909; Wolfe, 1986) (Blake, this volumeXX) make clear the disconnect between what is on the retina and what is perceived. It is also worth noting that, while acuity and other visual functions fall off rapidly with distance from fixation, our perception does not seem comparably degraded. Until we fixate and direct attention to the periphery, it is usually not obvious that very little of the visual field contains well-focused detail. This suggests that what we see can be built up over multiple fixations (e.g Noton & Stark, 1971) or deployments of attention even if precise transaccadic integration cannot be found. On the other hand, it is obvious, but worth noting, that we don't just see some memory of the prior objects of attention. Closing the eyes fundamentally changes the experience. The experience of seeing is ultimately based on visual input. When it is not, we call it a dream or a hallucination.

There is a richness to visual experience that is at odds with the data reviewed above. We attend to one or, perhaps, a few objects at a time. Yet the visual world appears to contain many objects. Even in a brief exposure, too brief to permit eye movements, we seem to see a complete visual field filled with the qualities of visual experience at all points. Where does this richness come from?

There are at least three potential sources of perceptual richness. First, even if we can only attend to a single object at a time, this does not mean that the consequences of attention are entirely lost when attention moves elsewhere (no matter what I have previously argued Wolfe, 2000; Wolfe et al., 2000). Returning to Figure 7, it seems possible that successive deployments of attention might serve to populate the representation with objects. If you want to check if an object in the representation still corresponds with an object in the stimulus or, perhaps, if you want to identify a specific object in the representation, you would need to pass through an attentional bottleneck, However, multiple objects of attention might persist (We could call this "persistence" but the term is already used in a more basic sense, e.g. DiLollo, Lark, & Hogben, 1988 ; Francis, Grossberg, & Mingolla, 1994). One way to think about this possibility is to ask what it means to say that only one object can be attended <u>at a time.</u> While only one object may be attended at a specific instant, you do not experience the present time as an instant. As James (1890) and many others recognized, the "psychological" (or "sensible") present has a duration. Various methods of measuring this duration yield estimates of a "present" that is100s of msec in length. If we assume that attention can be deployed at a rate of 20-40 Hz (for a discussion see Moore & Wolfe, 2000), this would yield a perceptual experience of many objects even if only one were actually attended in the physical, instantaneous present.

The second source of a rich and spatially continuous perception is the preattentive visual information discussed at the start of this chapter. It seems clear that we are consciously aware of visual "stuff" (Adelson & Bergen, 1991) throughout the visual field without the need for that "stuff" to be selected by attention. It might be a good idea to consider this awareness of some visual input across the field to be evidence of diffuse attention since there are circumstances of attentional tunnel vision where this awareness seems to be lost (Williams, 1985). Whether it is diffuse attention or preattention, under normal circumstances, we seem to be aware of something like the texture of preattentive basic features throughout the visual field. As with the apparent multiplicity of objects, one must be cautious about declaring that subjects see this preattentive "stuff" because attention is required any time one wishes to get a subject to make an explicit response to the presence of a stimulus. Thus, subjects might see color at all locations but they can only respond to color, one object or location at a time. Implicit measures can show that features were registered (e.g. Houck & Hoffman, 1986) but they cannot demonstrate perception. The fact that the representation (if any) is walled off from stimulus and/or from response by attentional bottlenecks seems likely to render the experimental evidence forever ambiguous. Thus, even if you are convinced that you have a rich perceptual life and even if you are willing to assume that others do too, it will be difficult, if not impossible to prove the point.

A third factor that could contribute to the creation of a rich perceptual representation is the ability to extract some meaning from unbound, minimally attended stimuli. It is striking that people appear to be able to gain some understanding of the meaning of a scene very quickly (Intraub, 1980; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001). The times required to show evidence of semantic processing are short enough that is difficult to imagine that the meaning is extracted by a succession of attentional deployments to a succession of objects. Recently, Oliva and Torralba (2001) have shown that it possible, in principle, to extract meaning from nonlocalized structural information encoded in the frequency spectrum. They devised a series of simple, linear, feed-forward filters that can be used to classify scenes on axes such as natural-artificial, rough-smooth, and open-closed. In the space defined by these axes, scenes naturally cluster into semantically meaningful categories like mountain scenes, beach scenes, street scenes, and so forth. We could call this the *"unbound semantics"* of the image.

Given these sources of information it should be possible for an observer to quickly develop a theory about the stimulus. That theory will be modulated by the observer's current biases and predispositions, The idea that the contents of the observer's mind might make a difference is an old one - enshrined in Shakespeare ("In the night, imagining some fear, how easy is a bush supposed a bear" Midsummer Night's Dream 5:1:21-22). In vision research search it is perhaps best known in Helmholtz's

(1924)discussion of "unconscious inference". Bayesian theories are the modern incarnation of this thought (Brainard & Freeman, 1997; Freeman, 1994; Lee, 1995). In all of these varying degrees of sophistication, the core idea is that what we experience as seeing is a theory. Work on attention adds to this idea by emphasizing the tenuous nature of the link between the theory that we see and the stimulus on the retina.

These ideas can be summarized by referring to another cartoon.

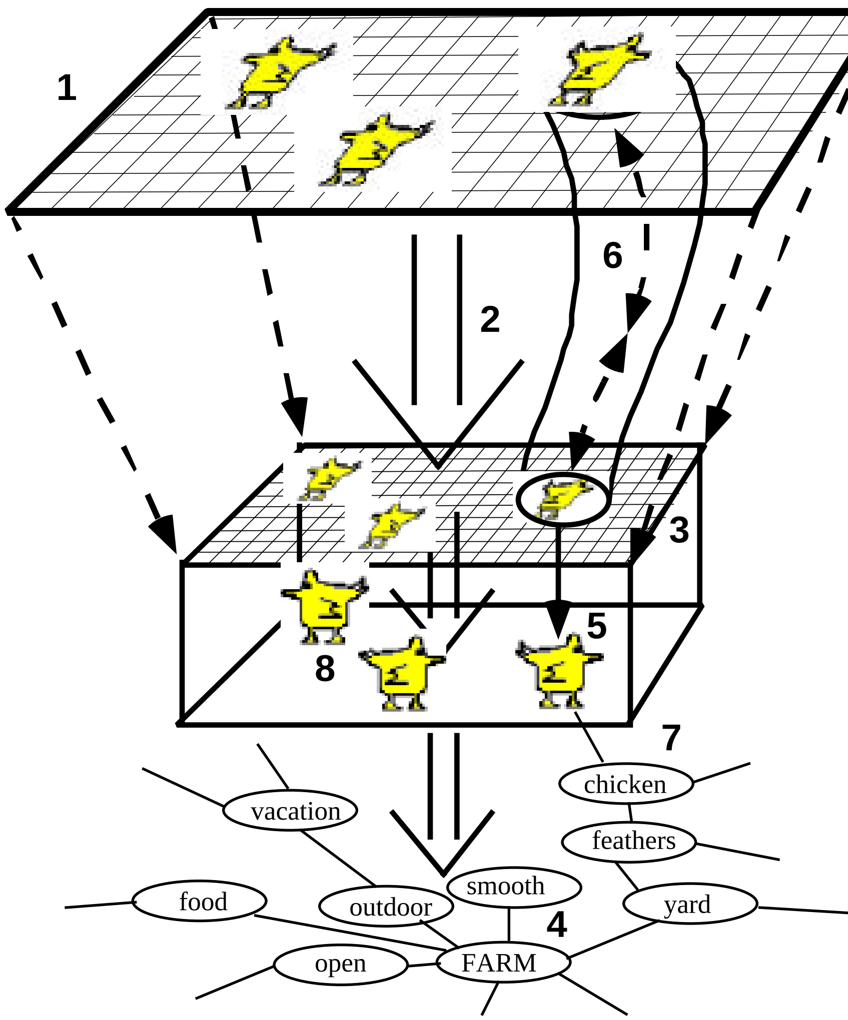

Figure Eight: A cartoon illustrating how the stimulus (1) might be processed in a massively parallel fashion (2) to yield a coarse, pre-attentive representation of the world (3). Further parallel processing might yield the unbound semantic notion that this was a farm scene (4). Representation of individual attended objects (like the chicken - 5) can be bound via feedback to early levels (6) and linked to the relevant nodes in memory (7). Object recently attended may be post-attentively represented in perception (8) even though they are not currently linked to their counterparts in the stimulus (note the mismatch in orientation of the "real" and "seen" middle chicken).

Imagine that you are looking at a farm scene with a few chickens in the yard. The farm scene is projected onto your retina and fed forward to early vision levels of analysis (1). The box at the center of the cartoon is the hypothetical level of the perceptual representation. Massively parallel, feed-forward pathways (2) create the coarse preattentive aspects of perception (3) and provide the data for the unbound semantic analysis that might identify this as a farm scene (4). In the cartoon, attention is directed to one of the chickens (5). This puts the perceived chicken in contact with the stimulus (6) and makes it possible to link the chicken to the relevant node in memory (7). Other, previously attended chickens (8) remain part of the perceptual experience even though the perceived chickens are not in current contact with the stimulus. Indeed, in the cartoon, the second chicken is facing different directions in the "world" and in the perceptual representation. This mismatch would not be noted or corrected until attention was redeployed to the relevant chicken.

Conclusions

It should be obvious that this paper has moved from the concrete to the speculative. There are some things that we know.

1) Some features are processed in parallel across the entire visual field
2) The resulting preattentive representation of these features is coarse, perhaps categorical. It is cruder than the analysis that is performed by early visual processing stages (e.g. primary visual cortex).
3) Preattentive vision appears to parse the input into candidate objects.
4) In the absence of attention, features appear to be only loosely bundled with their object.
5) Attention can be directed to those objects. This permits more accurate binding of features to objects.

6) Attention permits recovery of detailed information that is not available preattentively. This may involve feedback / re-entrant pathways.

7) Attention permits linkage between an object and its representation in memory, thus allowing the object to be recognized.

8) Only one (or perhaps a few) objects can be attended at one time.

9) If attention has been deployed elsewhere, it must be redirected back to an object in order to confirm that object's presence (e.g. in repeated search tasks) or to detect a change (e.g. change blindness).

10) Finally, these facts must somehow be reconcilable with the overarching fact that observers think that they are perceiving a visual world that extends across the visual field and is richly populated with coherent objects.

The proposal sketched above tries to reconcile #1-9 with #10 by accepting the old idea that we see our current "theory" about the external world and adding the notion that this theory is in very limited contact with the world at any moment in time.

Acknowledgements

I thank Jennifer DiMase and Aude Oliva for comments on drafts of this chapter. This research was supported by grants from the National Institutes of Health: National Eye Institute and National Institute for Mental Health, National Science Foundation, and the Air Force Office of Scientific Research.

REFERENCES

Adelson, E. H., & Bergen, J. R. (1991). The plenoptic function and the elements of early vision. In M. Landy & J. A. Movshon (Eds.), *Computational Models of Visual Processing.* (pp. 3-20). Cambridge, MA: MIT Press.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and visual hierarchy: Counter-streams in sensory processing and perceptual learning. *Nature, 387*(22 May 1997), 401-406.

Aks, D. J., & Enns, J. T. (1996). Visual search for size is influenced by a background texture gradient. *J. Experimental Psychology: Human Perception and Performance, 22*(6), 1467-1481.

Ballard, D., Hayhoe, M., & Pelz, J. (1995). Memory Representations in Natural Tasks. *J. Cognitive Neuroscience, 7*, 66-80.

Ballard, D., Hayhoe, M., & Pook, P. (1995). *Deictic Codes for the Embodiment of Cognition* (TR95.1): U of Rochester.

Barlow, H. (1995). The neuron doctrine in perception. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 415-435). Cambridge, MA: MIT Press.

Bauer, B., Jolicoeur, P., & Cowan, W. B. (1996). Distractor heterogeneity versus linear separability in colour visual search. *Perception, 25*(1), 1281-1294.

Bauer, B., Jolicoeur, P., & Cowan, W. B. (1998). The linear separability effect in color visual search: Ruling out the additive color hypothesis. *Perception and Psychophysics, 60*(6), 1083-1093.

Bauer, B., Jolicœur, P., & Cowan, W. B. (1996). Visual search for colour targets that are or are not linearly-separable from distractors. *Vision Research, 36*(10), 1439-1466.

Baylis, G. C., & Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *J. Exp. Psychol. - Human Perception and Performance, 19*(3), 451-470.

Beck, J. (1966). Perceptual grouping produced by changes in orientation and shape. *Science, 154*, 538-540.

Blackmore, S. J., Brelstaff, G., Nelson, K., & Troscianko, T. (1995). Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes. *Perception, 24*, 1075-1081.

Blake, R. (1989). A neural theory of binocular rivalry. *Psych. Review, 96*, 145-167.

Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. *J Opt Soc Am A, 14*(7), 1393-1411.

Brainard, D. H., Wandell, B. A., & Chichilnisky, E.-J. (1993). Color constancy: From physics to appearance. *Current Directions in Psychological Science, 2*(5), 165-170.

Braun, J. (1993). Shape-from-shading is independent of visual attention and may be a texton. *Spatial Vision, 7*(4), 311-322.

Bravo, M., & Blake, R. (1990). Preattentive vision and perceptual groups. *Perception, 19,* 515-522.

Breese, B. B. (1909). Binocular Rivalry. *Psych. Rev., 16*, 410-415.

Carlson-Radvansky, L. A. (1999). Memory for relational information across eye movements. *Perception and Psychophysics, 61*(5), 919-934.

Carter, R. C. (1982). Visual search with color. *J. Experimental Psychology: Human Perception and Performance, 8(1)*, 127- 136.

Cavanagh, P., Arguin, M., & Treisman, A. (1990). Effect of surface medium on visual search for orientation and size features. *J. Experimental Psychology: Human Perception and Performance, 16*(3), 479-492.

Cohen, A., & Ivry, R. B. (1989). Illusory conjunction inside and outside the focus of attention. *J. Exp. Psychol. Human Perception and Performance, 15*, 650-663.

Cohen, A., & Ivry, R. B. (1991). Density effects in conjunction search: Evidence for coarse location mechanism of feature integration. *J. Exp. Psychol. Human Perception and Performance, 17*(4), 891-901.

D'Zmura, M. (1991). Color in visual search. *Vision Research, 31*(6), 951-966.

Di Lollo, V., Enns, J. T., & Rensink, R. A., ,. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129*, 481-507.

DiLollo, V., Lark, C. D., & Hogben, J. H. (1988). Separating visible persistence from retinal afterimages. *Perception and Psychophysics, 44*, p363- 368.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*, 433-458.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting attention between objects and loctions: Evidence from normal and parietal lesion subjects. *J. Experimental Psychology: General*, *123*, 161-177.

Elder, J., & Zucker, S. (1993). The effect of contour closure on the rapid discrimination of two-dimensional shapes. *Vision Research, 33*(7), 981-991.

Enns, J. T., & Kingstone, A. (1995). Access to global and local properties in visual search for compound stimuli. *Psychological Science, 6*(5), 283-291.

Enns, J. T., & Rensink, R. A. (1990a). Scene based properties influence visual search. *Science, 247*, 721-723.

Enns, J. T., & Rensink, R. A. (1990b). Sensitivity to three-dimensional orientation in visual search. *Psychological Science, 1*(5), 323-326.

Foster, D. H., & Ward, P. A. (1991). Asymmetries in oriented-line detection indicate two orthogonal filters in early vision. *Proceedings of the Royal Society (London B), 243*, 75-81.

Foster, D. H., & Westland, S. (1992). Fine structure in the orientation threshold function for preattentive line-target detection. *Perception, 22*(supplement 2 (ECVP - Pisa)), 6.

Foster, D. H., & Westland, S. (1998). Multiple groups of orientation-selective visual mechanisms underlying rapid oriented-line detection. *Proc. R. Soc. Lond. B, 265*, 1605-1613.

Francis, G., Grossberg, S., & Mingolla, E. (1994). Cortical dynamics of feature binding and reset: Control of visual persistence. *Vision Research, 34*(8), 1089-1104.

Freeman, W. T. (1994). The generic viewpoint assumption in a framework for visual perception. *Nature, 368*(7 APril 1994), 542-545.

Gilchrist, I. D., W., H. G., Jane, R. M., & Heiko, N. (1997). Luminance and edge information in grouping: A study using visual search. *Journal of Experimental Psychology: Human Perception & Performance, 23*(2), 464-480.

Goldsmith, M. (1998). What's in a location? Comparing object-based and space-based models of feature integration in visual search. *J. Experimental Psychology: General, 127*(2), 189-219.

Green, B. F., & Anderson, L. K. (1956). Color coding in a visual search task. *J. Exp. Psychol.*, *51*, 19-24.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Perception* (pp. 89-110). New York: Oxford University Press.

Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research, 38*(1), 125-137.

Helmholtz, H. v. (1924). *Treatise on Physiological Optics* (Southall, Trans. Trans. from 3rd German ed. of 1909, ed.). Rochester, NY: The Optical Society of America.

Henderson, J. M. (1997). Transsaccadic memory and integration during real-world object perception. *Psychological Science, 8*(1), 51-55.

Hollingworth, A., & Henderson, J. M. (2001). Accurate visual memory for previously attended objects in natural scenes. *J. Exp. Psychol: Human Perception and Performance, in press*.

Houck, M. R., & Hoffman, J. E. (1986). Conjunction of color and form without attention. Evidence from an orientation-contingent color aftereffect. *J. Exp. Psychol: Human Perception and Performance, 12*, 186-199.

Houck, M. R., & Hoffman, J. E. (1986). Conjunction of color and form without attention: Evidence from an orientation contingent color aftereffect. *JEP: HPP, 12*, p186-199.

Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *J. Experimental Psychology: Human Learning and Memory, 6*(1), 1-12.

Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science, 5*(3), 94-100.

Irwin, D. E., Yantis, S., & Jonides, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception and Psychophysics, 34*(1), 49-57.

Irwin, D. E., Zacks, J. L., & Brown, J. S. (1990). Visual memory and the perception of a stable visual environment. *Perception and Psychophysics, 47*, 35- 46.

James, W. (1890). *The Principles of Psychology*. New York: Henry Holt and Co.

Julesz, B., & Bergen, J. R. (1983). Textons, the fundamental elements in preattentive vision and perceptions of textures. *Bell Sys. Tech. J., 62*, 1619- 1646.

Karn, K., & Hayhoe, M. (2000). Memory representations guide targeting eye movements in a natural task. *Visual Cognition., 7*, 673-703.

Kellman, P. (1998). An update on gestalt psychology, *Perception, Cognition, and Language: Essays in honor of Henry and Lila Gleitman* (pp. ??). Cambridge, MA: MIT Press.

Kleffner, D., Polichar, V. E., & Ramachandran, V. S. (1990). Shape from shading affects motion perception and brightness constancy. *Invest. Ophthalmol. Vis. Sci. (ARVO suppl), 31*, 524.

Kristjansson, A., & Tse, P. U. (2001). Curvature discontinuities are cues for rapid shape analysis. *Perception and Psychophysics, 41*(3), 390-403.

Lamme, V. A. F., Super, H., & Spekresijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology, 8*, 529-535.

Lee, T. S. (1995). A bayesian framework for understanding texture segmentation in the primary visual cortex. *Vision Research, 35*(18), 2643-2657.

Lettvin, J. (1995). J Y Lettvin on grandmother cells. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 434-435). Cambridge, MA: MIT Press.

Logan, G. (1992). Attention and preattention in theories of automaticity. *American Journal of Psychology, 105*(2), 317-339.

Mack, A., & Rock, I. (1998a). *Inattentional Blindness*. Cambridge, MA: MIT Press.

Mack, A., & Rock, I. (1998b). Inattentional blindness: Perception without attention. In R. D. Wright (Ed.), *Visual Attention* (Vol. 8, pp. 55-76). New York: Oxford.

Moore, C. M., & Wolfe, J. M. (2000). Getting beyond the serial/parallel debate in visual search: A hybrid approach. In K. Shapiro (Ed.), *The Limits of Attention: Temporal Constraints on Human Information Processing*. Oxford: Oxford U. Press.

Moraglia, G. (1989). Display organization and the detection of horizontal lines segments. *Perception and Psychophysics, 45*, 265-272.

Nagy, A. L., & Sanchez, R. R. (1990). Critical color differences determined with a visual search task. *J. Optical Society of America - A, 7*(7), 1209-1217.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science, 257*(4 Sept), 1357-1363.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton, Century, Crofts.

Nelson, J. I. (1974). Motion sensitivity in peripheral vision. *Perception*, *3*, 151-152.

Neumann, E., & DeSchepper, B. G. (1992). An inhibition-based fan effect: Evidence for an active suppression mechanism in selective attention. *Canadian Journal of Psychology, 46*, 1-40.

Nothdurft, H. C. (1992). Saliency in visual search. *Science, MS REVIEWED 3.5.92 DO NOT CITE*.

Noton, D., & Stark, L. (1971). Eye movements and visual perception. In R. Held & W. Richards (Eds.), *Perception: Mechanisms and Models - Readings from Scientific American* (pp. 218-227). San Francisco: WH Freeman.

O'Connell, K. M., & Treisman, A. M. (1990). Is all orientation created equal? *Investigative Ophthalmology and Visual Science, 31*(4), 106.

O'Regan, J. K., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*.

O'Regan, K. (1992). Solving the 'real' mysteries of visual perception. The world as an outside memory. *Canadian J. of Psychology, 46*, 461-488.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145-175.

Previc, F. H., & Naegele, P. D. (2001). Target-tilt and vertical-hemifield asymmetries in free-scan search for 3-D targets. *Perception and Psychophysics, 41*(3), 445-457.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature, 331*, 163-165.

Rensink, R., & Cavanagh, P. (1993). Processing of shadows at preattentive levels. *Investigative Ophthalmology and Visual Science, 34*(4), 1288.

Rensink, R., & Cavanagh, P. (1994). Identification of highlights in early vision. *Investigative Ophthalmology and Visual Science, 35*(4), 1623.

Rensink, R., O'Regan, J. K., & Clark, J. J. (1996). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*, 368-373.

Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vision Res, 40*(10-12), 1469-1487.

Rensink, R. A., & Enns, J. T. (1995). Pre-emption effects in visual search: evidence for low-level grouping. *Psychological Review, 102*(1), 101-130.

Rock, I., & Palmer, S. (1990). The legacy of Gestalt psychology. *Sci Am, 263*(6), 84-90.

Roelfsema, P. R., Lammer, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395*, 376.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*(7), 261-267.

Simons, D. J. e. (2000). *Change blindness and visual memory.*

Sun, J., & Perona, P. (1996a). Early computation of shape and reflectance in the visual system. *Nature, 379*(11 Jan), 165-168.

Sun, J., & Perona, P. (1996b). Where is the sun? *Investigative Ophthalmology and Visual Science, 37*(3), S935.

Theeuwes, J., & Kooi, J. L. (1994). Parallel search for a conjunction of shape and contrast polarity. *Vision Research, 34*(22), 3013-3016.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6 June), 520-552.

Tipper, S. P., & Weaver, B. (1998). The medium of attention: Location-based, object-based, or scene-based? In R. D. Wright (Ed.), *Visual Attention* (Vol. Vol 8, pp. 77-107). Oxford: Oxford U. Press.

Tipper, S. P., Weaver, B., Jerreat, L. M., & Burak, A. L. (1994). Object-based and environment-based inhibition of return of visual attention. *Journal of Experimental Psychology: Human Perception & Performance, 20*(3), 478-499.

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.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology, 6*, 171-178.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psych. Review, 95*, 15-48.

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of seperable features. *J. Exp. Psychol. - General, 114*, 285-310.

Treisman, A. M., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psych., 14*, 107-141.

Tsotsos, J. K. (1988). A 'complexity level' analysis of immediate vision. *International J of Computer Vision*, p303-320.

Tsotsos, J. K. (1993). An inhibitory beam for attentional selection. In L. Harris & M. Jenkin (Eds.), *Spatial Vision in Humans and Robots* (pp. 313 - 331). Cambridge, UK: Cambridge U. Press.

Tsotsos, J. K., Culhane, S. N., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence, 78*, 507-545.

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and

artifactual objects. *Perception, 30*(6), 655-668.

Williams, L. G. (1966). The effect of target specification on objects fixed during visual search. *Perception and Psychophysics, 1*, 315-318.

Williams, L. J. (1985). Tunnel vision induced by a foveal load manipulation. *Human Factors, 27*(2), 221-227.

Wolfe, J. (2000, April 10-15, 2000). *Post-attentive vision and the illusion of perception.* Paper presented at the Toward a Science of Consciousness, Tuscon, AZ.

Wolfe, J. M. (1986). Stereopsis and binocular rivalry. *Psych. Review, 93*, 269-282.

Wolfe, J. M. (1992). "Effortless" texture segmentation and "parallel" visual search are *not* the same thing. *Vision Research, 32*(4), 757-763.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13-74). Hove, East Sussex, UK: Psychology Press Ltd.

Wolfe, J. M. (2001). Asymmetries in Visual Search: An Introduction. *Perception and Psychophysics, 63*(3), 381-389.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive Object Files: Shapeless bundles of basic features. *Vision Research, 37*(1), 25-44.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided Search: An alternative to the Feature Integration model for visual search. *J. Exp. Psychol. - Human Perception and Perf., 15*, 419-433.

Wolfe, J. M., & Franzel, S. L. (1988). Binocularity and visual search. *Perception and Psychophysics, 44*, 81-93.

Wolfe, J. M., & Friedman-Hill, S. R. (1992). On the role of symmetry in visual search. *Psychological Science, 3*(3), 194-198.

Wolfe, J. M., & Friedman-Hill, S. R. (1992). Visual search for orientation: The role of angular relations between targets and distractors. *Spatial Vision, 6*(3), 199-208.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *J. Exp. Psychol: Human Perception and Performance, 18*(1), 34-49.

Wolfe, J. M., Klempen, N., & Dahlen, K. (2000). Post-attentive vision. *Journal of Experimental Psychology:Human Perception & Performance, 26*(2), 693-716.

Wolfe, J. M., Yee, A., & Friedman-Hill, S. R. (1992). Curvature is a basic feature for visual search. *Perception, 21*, 465-480.

Yantis, S., & Hillstrom, A. P. (1994). Stimulus-driven attentional capture: Evidence from equiluminant visual objects. *J. Experimental Psychology: Human Perception and Performance, 20*(1), 95-107.

Yantis, S., & Jonides, J. (1996). Attentional capture by abrupt onsets: New perceptual objects or visual masking. *J. Experimental Psychology: Human Perception and Performance, 22*(6), 1505-1513.