# View from the Top: Hierarchies and Reverse Hierarchies in the Visual System

**Review**

**Shaul Hochstein[1,3,4] and Merav Ahissar[2,3]**
[1]Department of Neurobiology
[2]Department of Psychology
[3]Neural Computation Center
Hebrew University
Jerusalem, 91904
Israel

We propose that explicit vision advances in reverse hierarchical direction, as shown for perceptual learning. Processing along the feedforward hierarchy of areas, leading to increasingly complex representations, is automatic and implicit, while conscious perception begins at the hierarchy's top, gradually returning downward as needed. Thus, our initial conscious percept—*vision at a glance*—matches a high-level, generalized, categorical scene interpretation, identifying "forest before trees." For later *vision with scrutiny*, reverse hierarchy routines focus attention to specific, active, low-level units, incorporating into conscious perception detailed information available there. Reverse Hierarchy Theory dissociates between early explicit perception and implicit low-level vision, explaining a variety of phenomena. Feature search "pop-out" is attributed to high areas, where large receptive fields underlie spread attention detecting categorical differences. Search for conjunctions or fine discriminations depends on reentry to low-level specific receptive fields using serial focused attention, consistent with recently reported primary visual cortex effects.

*Life can only be understood backward, but it must be lived forward.*

—Søren Kierkegaard

## Introduction

Central to cognitive neuroscience today are three puzzles related to two modes of perception and neuronal processing. These related challenges concern the apparent dichotomies between (1) rapid yet generalized versus slower detailed perception, (2) between implicit and often automatic versus conscious and attentive perception, and (3) between perceptual learning effects that transfer to new stimulation conditions and those that are specific to the conditions of original training. We will show that these issues may be related to each other and to the division between feedforward, perhaps automatic, visual processing versus mechanisms that are driven by focused attention and depend on the massive feedback connections found along the visual cortical hierarchy.

(1) The first issue is the apparent disparity between our extremely rapid capture of the conceptual gist of a scene together with our blindness to its details. Since the whole is surely built of its parts, how is it that the parts remain unknown, while the whole becomes accessible? This perceptual enigma is exemplified in perception under Rapid Serial Visual Presentation (an experimental paradigm called RSVP, where pictures are presented in rapid sequence, typically 10 to 16 per second, and observers are asked to detect and report something about one or all of them) and in related repetition and change blindnesses (the failure of observers to report detail changes in alternating pictures or the presence of two identical or similar objects). At ten images per second, observers can identify picture or word categories (Potter, 1976) but are insensitive to repetitions of the same object (despite differences in size, orientation, or viewpoint; Kanwisher, 1987). Observers also fail to see large changes in a visual scene, even at slower rates, as long as the gist remains the same (Rensink et al., 1997). The phenomenon of repetition blindness seems especially paradoxical: it implies that processing has proceeded to the level of determining that one element is a conceptual or categorical repetition of another—a repetition to which we are then blind. How can we know that two elements are similar if we are blind to the double occurrence? A similar paradox appears when we briefly view a scene containing many elements. We can more easily report the average value of a parameter (such as the mean size or orientation of elements) than judge if a particular value was present in the scene (Ariely, 2001). How is the average computed without knowledge of the exemplars? Our answer to these questions will be that *implicit* processes make use of basic information (number of similar elements, detailed parameters of each element) to derive resulting higher-level percepts (categories represented and their average values but not their number or precise location). However, only the results of these computations are accessible to conscious perception.

(2) The second issue concerns the very nature of rapid perception, addressed cogently with the visual search paradigm. Two search modes have been discerned: rapid, easy, feature search, which appears to be conducted automatically and in parallel across an array of elements, compared to slower, presumably serial search, which requires focused attention and shows a definite set-size effect (Treisman and Gelade, 1980). Following more than two decades of research, it remains an open question what a priori rules define the mode of a particular search. Though rapid search was originally found for quite simple features (suggesting to some a low-level underlying site), it was subsequently also found for quite high-level features, which include the results of complex processing procedures (see below). On the other hand, even the search for simple features is performed serially when the difference between target and distractors is small—though well above the discrimination threshold. What then determines whether visual search will be automatic and parallel or attentive and serial? We will attribute the rapid process to postimplicit processing levels.

(3) The third question relates to the spread of training
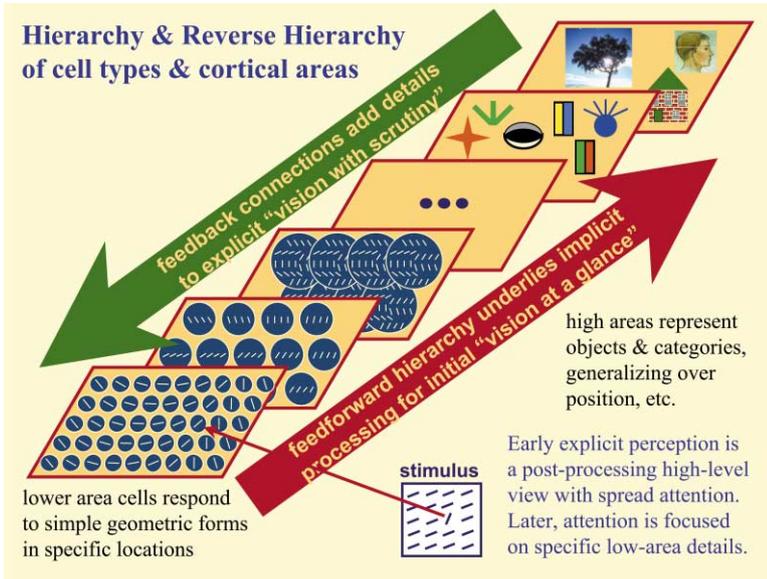
[4]Correspondence: shaul@vms.huji.ac.il

Figure 1. Schematic Diagram of Classical Hierarchy and Reverse Hierarchy Theory

Classically, the visual system was seen as a hierarchy of cortical areas and cell types. Neurons of low-level areas (V1, V2) receive visual input and represent simple features such as lines or edges of specific orientation and location. Their outputs are integrated and processed by successive cortical levels (V3, V4, medial-temporal area MT), which gradually generalize over spatial parameters and specialize to represent global features. Finally, further levels (inferotemporal area IT, prefrontal area PF, etc.) integrate their outputs to represent abstract forms, objects, and categories. The function of feedback connections was unknown. Reverse Hierarchy Theory proposes that the above forward hierarchy acts implicitly, with explicit perception beginning at high-level cortex, representing the *gist of the scene* on the basis of a first-order approximate integration of low-level input. Later, explicit perception returns to lower areas via the feedback connections, to integrate into conscious *vision with scrutiny* the detailed information available there. Thus, initial perception is based on spread attention (large receptive fields), guessing at details, and making binding or conjunction errors. Later vision incorporates details, overcoming such blindnesses.

effects. There are cases where training with one set of spatial stimulus conditions affects later performance of the same task under somewhat different spatial stimulus conditions. In contrast, for apparently similar cases, learning effects may be substantially specific to the conditions of training (reviewed in Ahissar and Hochstein, 2002). What determines the extent of transfer of learning effects? And when can we expect training to generalize to new stimulation conditions? Recent results suggest that early, easy-case learning is generalized, while slower, difficult-case learning may be specific to spatial conditions (Ahissar and Hochstein, 1997, 2000). We suggest that the degree of generalization depends on the receptive field characteristics of neurons underlying the training effects. (The receptive field of a single neuron is the collection of visual stimuli that affect its response, including their locations and spatial characteristics. For example, a V1 neuron might respond to a bar of light that is narrower than $1°$ of visual angle and shorter than $10°$, tilted at $30°–50°$ clockwise from the vertical, and located at a specific region of the visual field. A neuron in inferotemporal [IT] cortex may respond to a specific face of size between $4°–12°$ and location between $-2°$ to $+10°$ from the vertical meridian, viewed at a variety of angles from frontoparallel and from vertical.)

The key to resolving these three dichotomies may lie in the hierarchy of processing levels in the visual system and the feedforward and massive feedback connections among them. The functional roles and cognitive consequences of this physiological-anatomical organization are still unknown. In particular, the sites of attention and conscious perception have not been identified. Relating to the above dichotomies, we wish to pinpoint (1) where in the hierarchy are the mechanisms underlying very rapid perception of the gist of a scene and where are those underlying slower access to its details; (2) where are the procedures supporting parallel and where are those for serial search modes; and (3) where are the

sites of generalized and where are those of specific learning?

## Classical View of the Visual Hierarchy

Classical physiological studies of visual cortex converged on three underlying principles. First, processing as revealed by neuronal receptive fields is hierarchical, beginning with the Simple and Complex cells described by Hubel and Wiesel (1968). Neurons in higher cortical areas represent increasingly complex characteristics of images, objects, and categories (Maunsell and Newsome, 1987; Vogels and Orban, 1996). Higher neurons are also less dependent on coincidental spatial features, such as precise location, retinal size, viewpoint, lighting, and color. This hierarchy is schematically represented in Figure 1. The second feature of the classical view is the presence of multiple areas belonging to parallel pathways, such as the dorsal and ventral streams (Ungerleider and Mishkin, 1982; Livingstone and Hubel, 1988; Goodale and Milner, 1992).

The third principle, that short- and long-term cortical plasticity are limited to particular cortical sites, has undergone substantial recent revision. Until the last decade, learning (i.e., experience) and attention (i.e., context) were seen as influencing high cortical areas (inferotemporal cortex or its human analog the lateral-occipital complex and perhaps area V4 but not V1; Bushnell et al., 1981; Moran and Desimone, 1985; Malach et al., 1995). Although it was clear that feedback connections reach as low as V1 and even the LGN, their function was largely ignored. As discussed below, recent physiological and functional MRI studies have now confirmed top-down effects even in V1.

## Reverse Hierarchy of Perceptual Learning

We performed a series of studies on perceptual learning and found that even training very simple tasks leads to considerable performance improvement—seen as an
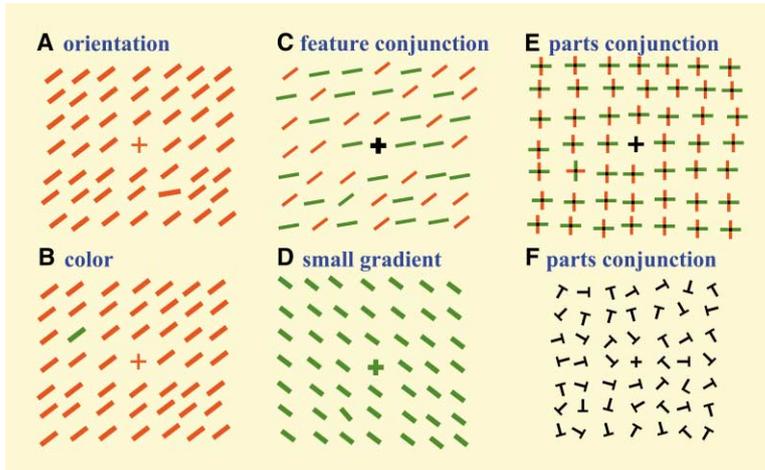
Figure 2. Feature Integration Theory

Odd elements "pop-out" when they differ substantially from the distractors in one feature, such as orientation (A) or color (B). Search with serial focused attention is required if the odd element differs only in the conjunction of features (C) or the spatial arrangement of element parts (E and F) or if the difference is small (D) (see Treisman and Gelade, 1980; Treisman, 1988). Reverse Hierarchy Theory assigns the pop-out phenomenon to initial perception at high-level areas using their large receptive fields, i.e., spread attention. Later feedback reentry to low levels slowly adds details available in the small specific receptive fields found there.

increase of up to a factor three in processing speed (Ahissar and Hochstein, 1993). The task used was orientation pop-out (Figure 2A). Subjects were consistently trained with one set of conditions and subsequently tested with another set (e.g., with different target and/ or distractor element orientations, sizes, or locations). Sometimes the learning effects transferred to these new conditions, and sometimes they did not (reviewed in Ahissar and Hochstein, 2000, 2002; see bibliographies there and chapters in Fahle and Poggio, 2002, for similar findings with many other perceptual tasks). Transfer was measured with several levels of difficulty determined by target-distractor orientation difference, target position uncertainty, and stimulus-to-mask delay. Systematic analysis led to the conclusion that, as a rule, with easy task conditions there is considerable learning transfer, while with hard task conditions there is more learning specificity (Ahissar and Hochstein, 1997; Liu, 1999). Furthermore, there was a cascade of learning, such that easy conditions were learned early and hard conditions were learned later, as demonstrated in Figures 3A and 3B. This temporal order was mandatory. That is, when training was attempted with only difficult cases, typically, no improvement was found. Learning difficult cases only began after at least one clear and easy example had been given (Ahissar and Hochstein, 1997; Rubin et al., 1997).

An important contribution to our understanding came by seeing transfer as a diagnostic sign of the cortical level of learning (Fiorentini and Berardi, 1980; Ahissar and Hochstein, 1997). Learning transfer derives from modification at neurons with generalizing receptive fields, found at high cortical levels, and specificity derives from changes at neurons with localized receptive fields, found at low cortical levels. Thus, easy-condition learning transfer reflects tasks performed and learned at high cortical levels, while hard-condition specificity is a sign of tasks depending on low cortical level mechanisms. The cascade of learning from easy to hard conditions suggests the conclusion that high-level learning occurs before low-level learning, i.e., in reverse hierarchy order. The mandatory nature of this learning order derives from low cortical level (hard condition) learning needing guidance from prior (easy condition), high-level learning.

In summary, using transfer as a diagnostic sign of cortical level, evidence from perceptual learning experiments suggests the following. (1) Easy-condition learning occurs at high cortical levels, while difficult-condition training affects low levels. (2) There is a cascade from initial high-level to subsequent low-level learning. That is, learning occurs in the visual cortex in reverse
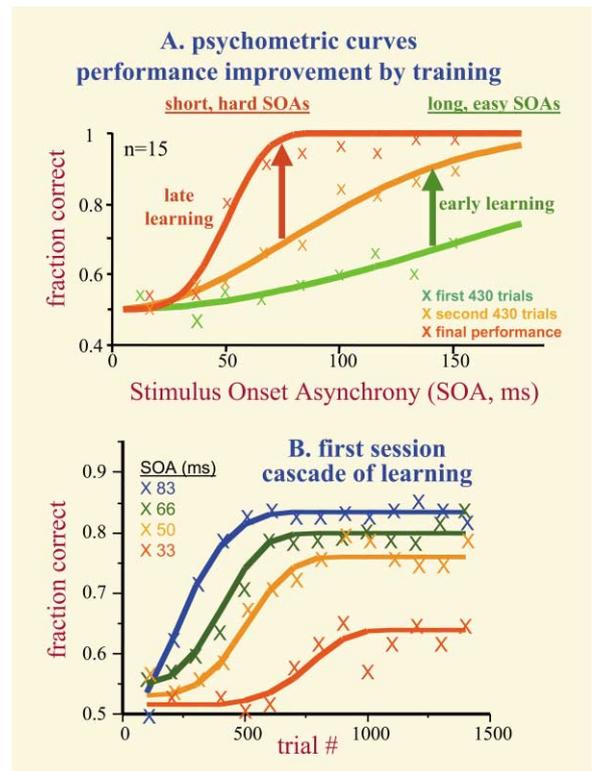


Figure 3. Cascade of Perceptual Learning

Training with interleaved easy and difficult trials (longer and shorter stimulus-to-mask onset asynchronies or SOAs) leads to initial learning of easy trials and subsequent learning of more difficult trials. This is seen in early changes in the long SOA part of the psychometric function and later changes in the short SOA part (A). Measuring performance for each SOA, learning effects are clearly seen to have different time courses for the different SOAs (B). (Adapted from Ahissar and Hochstein, 1997).

hierarchy order. (3) High-level training effects enable and guide later learning at lower levels. We interpret the inability of the system to directly initiate low-level learning as deriving from its limited access to appropriate populations at these levels. When nonsalient stimulus parameters need to be resolved, access is gained gradually, by higher-level node intervention.

We now suggest that the reverse hierarchy principles found for perceptual learning may also apply to conscious perception itself (though on a considerably shorter time scale). In this way, the reverse hierarchy sequence may be an inherent characteristic of visual perception.

## Reverse Hierarchy of Explicit Perception

We use the term "explicit perception" to include stimulus-driven experiences that are accessible to conscious identification and recognition (Goodale and Milner, 1992; Lumer and Rees, 1999) and/or to retrieval from long-term memory store (Squire, 1986). We assume that explicit perception is the reflection of cortical activity at one or more sites, and we wish to identify these sites. We do not assume that activity at a particular site always is, or is available to, the neural correlate of consciousness (Crick and Koch, 1995). For example, we take the stand that attention, whether focused or spread, is required for explicit perception.

We suggest a new view of visual perception, Reverse Hierarchy Theory (RHT), proposing that explicit visual perception follows the visual hierarchy in reverse direction, from top to bottom, as outlined in Figure 1. Initial feedforward implicit processing follows the bottom-up hierarchical pathway, but not all processing is directly available to conscious perception. Instead, explicit visual perception only begins when processing reaches high cortical levels and proceeds in top-down fashion to gradually encompass detailed information available at more peripheral cortical areas as needed. Details incorporated at later stages of conscious vision include precise location, retinal size and color, and component motion. The parallel forward and reverse hierarchy pathways mentioned above may be responsible for separate integration of form, motion, and color details. Our ostensibly unidimensional scheme does not directly address lateral interactions within and between cortical areas nor control functions of parietal or other areas. While these obviously play an essential role, our current focus will be the function of bottom-up and top-down hierarchical connections. The large receptive fields of high cortical areas are reflected in the spread attention of initial perception, while smaller low-area receptive fields are responsible for later focused attention. Initial explicit perception with spread attention "detects" objects by a first approximation "guess" as to the binding of features (arriving perhaps from separate streams) falling within the same large high-level receptive fields. Given this initial crude binding, one function of reentry is to confirm or refute such initial guesses.

Conscious vision appears as a continuum between two modes, related to representations in the reverse hierarchy top-down cascade of cortical areas. At the two extremes are *vision at a glance*, with spread attention depending on the large receptive fields of higher cortical areas, and *vision with scrutiny*, which incorporates details available only in the small receptive fields at lower cortical areas.

Under day-to-day conditions (rather than in controlled laboratory experiments), we survey scenes continuously rather than having pictures briefly flashed before us. Daily vision may still encompass the extreme *at a glance* and *with scrutiny* modes. The default procedure of explicit vision may include high-level perception with spread attention, supplemented by occasional forays to lower levels with focused attention. Nevertheless, we will generally speak in terms of the laboratory situation of initial versus later perception, which more easily separates these bottom-up (implicit) and top-down (explicit) processes.

## Characteristics of *Vision at a Glance*
### Initial Perception of Object Categories
As mentioned in the Introduction, the recently popular studies using RSVP (rapid serial visual presentation) demonstrated our phenomenal ability to rapidly detect basic-level categories (100–200 ms; Potter, 1976). For example, we can recall most words in a sentence presented at 12 words per second. With frame rates of 8 per second, we are able to detect the presence of a previewed picture, of a picture containing a pre-named category (e.g., an animal), or of a picture *not* containing the pre-named category (Intraub, 1999). The properties of this rapid detection match those of high-level cortical neurons representing basic categories. The high-level category neurons use detailed information implicitly— i.e., they receive input from low-level neurons whose receptive fields represent these details—but these details themselves are not represented in the high-level neurons and are therefore not immediately accessible to consciousness. Thus, it is easier to report a briefly presented word than its font or that a picture contained an apple than its color, even though the font characteristics or apple color were implicitly used to determine the categorical percept. The principle of high-level representation is that these categorical mechanisms generalize over space, size, viewpoint, etc., so that their activity indicates the presence of the basic category or object type but not its precise parameters.

While basic-level categories are detected at a glance, subordinate categories require further processing (Rosch et al., 1976). It may be thought that the latter processing occurs at the same or at even higher levels. Such higher levels have not been found, however. We suggest a third alternative: since subordinate category identification requires incorporation of fine details, it should involve reverse hierarchy routines. Thus, subordinate categories are determined by basic-level membership—represented by high-level neurons—supplemented by specific characteristics, represented lower. The difference between basic and subordinate categories is often a matter of experience with the environment. With training, new basic categories may be learned. Whether this learning derives from high-level neurons becoming tuned to the new categories is an open question.

### Blindnesses of Initial Perception
It has been found that together with the amazing ability to rapidly capture scene categories there is an equally
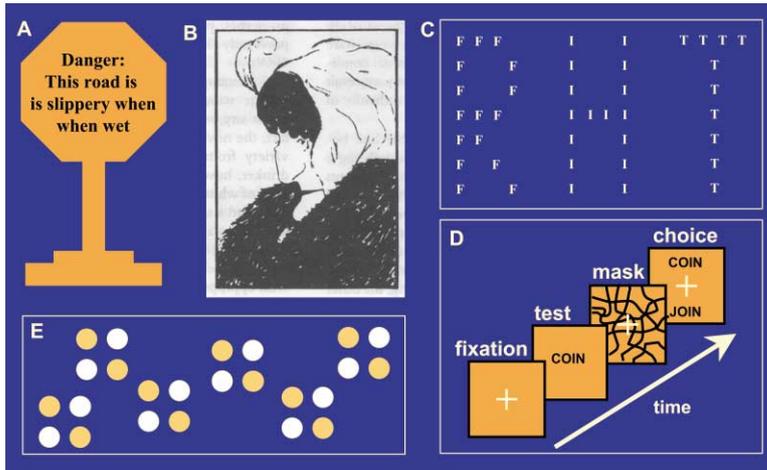
Figure 4. High Cortical Level Determination of Initial Conscious Visual Perception

(A) Repetition blindness. Categorical repetitions in time or space are often overlooked, as are the repeated words "is" and "when" in this sign (Kanwisher, 1987; Luo and Caramazza, 1996).

(B) The very first view of an ambiguous figure produces an integrated percept of one possible interpretation, not an ambiguous collection of lines and colored regions.

(C) When large letters are composed of small letters, the initial percept usually matches the more global stimulus (Navon, 1977).

(D) Subjects more easily detect a masked word than a masked letter, even if the difference between the words to be chosen from is only in that letter (Johnston and McClelland 1974).

(E) Apparent motion. If the image presented is made up of alternations between the white and the yellow dots, the global percept is of either horizontal or vertical motion of all the dots but not intermixed motions (Ramachandran and Anstis, 1983). Thus, initial explicit motion perception is global and high level.

amazing initial blindness to the details in a scene. For example, with RSVP, observers easily miss repetitions within closely successive pictures (or even in a single picture; Figure 4A), and they are often unable to report picture details. As pointed out above, an intriguing factor of repetition blindness is that the second presentation need not be an identical replica of the first, so that determining that it was a repetition is not at all trivial. Kanwisher (1987) suggested that we perceive "types" (categories) separately from "tokens" (specific exemplars), and in the case of RSVP, we pick up only the types. In our view, a brief exposure allows high-level cortical neurons to build initial representations of the gist of the scene, including generalized categories of elements within it but not their details. We thus interpret change blindness (where moderately large changes between successively viewed scenes are missed; Rensink et al., 1997) as another instance of the same initial-vision blindness (Figures 5 and 6). We predict that blindness will only be found for changes that do not affect the categories of the objects explicitly perceived in the scene. On the other hand, prolonged repeated viewing allows for the serial introduction of details, by returning to low-level cortical representations, with small receptive fields, focusing attention to one area or object at a time (Balz and Hock, 1997; Hock et al., 1998). Viewers only notice the changes when attention is serendipitously focused on the location of the change.

There are limits and costs, however, to including lower-level information with focused attention. The return to lower-area details is limited and serial in nature. At the same time, high-level *gist-of-the-scene* sensitivity is reduced when attention is focused rather than spread. These effects may be responsible for the phenomenon of inattentional blindness, whereby subjects often fail to report the appearance of an otherwise salient, distinctive, but unexpected item when they are in the midst of performing a task that demands focused attention (Mack and Rock, 1998; Most et al., 2001). Similarly, when attention is pre-cued to a particular location, the sensitivity of the high-level spread attention mechanism is

reduced—leading, for example, to decreased pop-out detection (Ahissar et al., 2001).

The lack of precision of default scene perception—when specific scrutiny of details is not called for—has also been demonstrated when observers are asked to draw a picture that they have seen. They tend to "remember" having seen a greater expanse of the scene than had been in the photograph (Intraub, 1999). These false-positive errors may be related to high-level processes perceiving object categories within the scene—and not (yet) the detail that these objects are only partially in view.

***Global Priority and Coherence of Initial Perception***

There are numerous other examples of the immediacy and priority of global perception. The Gestalt school
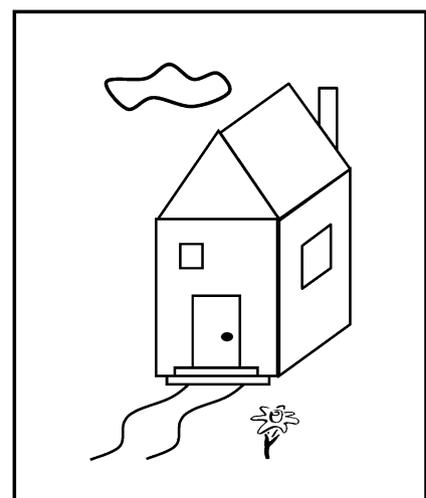


Figure 5. Change Blindness

Whereas high-level units rapidly acquire the *gist of the scene* by integrating bottom-up information to form a categorical representation, details are not acquired without later focusing attention to them. Observers often require multiple viewings to perceive most of the differences between this figure and Figure 6 on the following page.
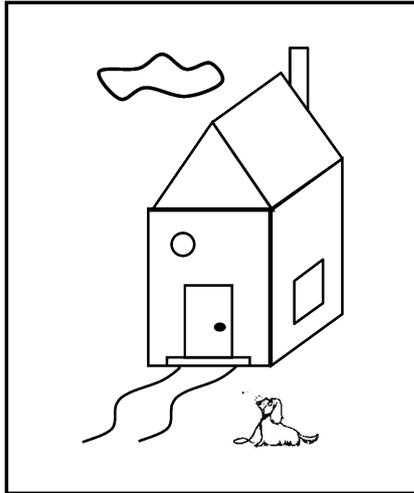
Figure 6. Change Blindness

Observers require "vision with scrutiny," serially focusing attention to different scene details, to find the differences between this figure and Figure 5 on the preceding page. Hint: besides the categorically and thus obviously different item on the lawn, look for positions or shapes of windows and the chimney, as well as the number of steps to the door.

emphasized the immediate perception of global attributes without conscious awareness that the visual system was processing individual elements to form a whole. According to the reverse hierarchy view, the whole is perceived first due to explicit perception initially accessing only high-level representations. Subsequent studies noted that we consciously perceive words before letters (Figure 4D; Johnston and McClelland, 1974), scenes before the objects composing them (Biederman et al., 1974), and in picturesque terms, "the forest before the trees" (Figure 4C; Navon, 1977). As mentioned above, subjects more easily report the average value of a parameter, such as the size or orientation of elements, in a scene than they can judge if a particular value was present in it (Ariely, 2001).

The initial explicit percept is of a complex coherent scene, even when it contains conflicting or ambiguous possible interpretations. For example, the immediate conscious percept of an ambiguous image (e.g., Figure 4B) is of one alternative based on global features, never an ambiguous sensation between or combining the two alternatives. Similarly, conflicting stimuli to the two eyes, yielding binocular rivalry, often produce a single (perhaps changing) percept. In the same way, conflicting aspects of a single picture may be integrated without noticing the conflicts, as in the "impossible" figures made famous by Escher. The initially perceived solution to the apparent motion correspondence problem is also a global one. Thus, in a multi-spot ambiguous motion situation, one may perceive the spots as moving in one or another direction, but one always perceives all the spots as moving in the same direction (Figure 4E; Ramachandran and Anstis, 1983). Similarly, plaids may be seen as moving in coherence or as gratings sliding over each other, but the earliest percept is of one or the other, never of ambiguously both. Again, the percept is

high level, as evidenced by its being influenced by color, global interactions, and adaptation or learning effects (Kooi et al., 1992). All of these are high-level aspects of initial explicit perception pointing to its following intricate implicit hierarchical processing. Low-level mechanisms can represent conflicting evidence in parallel (Cumming and Parker, 1997)—leading to adaptation and priming even for aspects not incorporated in the explicit percept (Hock et al., 1996). High-level representations, however, are generally mutually exclusive, leading to a single, globally constrained percept.

One advantage of reverse hierarchy processing may be that initial high area organization of information can be used for guiding attention to selected subgroups of previously activated low-level neurons (Cave and Wolfe, 1990). Top-down effects following initial feedforward processing allow for space- and/or object-based attention (Duncan, 1984; Vecera and Farah, 1994) and even attention following figure/ground determination (Wong and Weisstein, 1982). Reverse hierarchy guidance may follow the route of initially activated (feedforward) connections, focusing attention on appropriate active inputs to higher-level units (perhaps directed by parietal cortex and/or the Pulvinar, especially when multiple objects need to be attended; see Humphreys, 1998).

### Characteristics of Initial Feature Binding

There is another reason for associating initial aspects of explicit visual perception with high area detection mechanisms. If initial perception would depend only on separate low area feature maps, then initial vision should perceive colors, orientations, and motions, not conjoined forms. Yet, our very first *vision at a glance* is not of such unattached free-floating features but rather of coherent conjoined objects. The occasional presence of illusory conjunctions—i.e., incorrectly bound features, e.g., when associating with a presented shape the color of a neighboring shape—is due to incorrectly bound features, but bound features nevertheless. Such false conjunctions across retinal positions cannot be the outcome of locally restricted low-level visual representations. They must result instead from initial bottom-up convergence from multiple locations. The likelihood of false conjunctions in a manner that matches prior knowledge of the environment, such as round green apple and elongated yellow banana, is higher than that of other false conjunctions (Treisman, 1986). These biases reflect the initial assignment of familiar combinations of features to specific objects. Thus, initial object recognition incorporates a priori "assumptions" influenced by experience. These features of initial high-level vision are a natural and direct outcome of the receptive field properties of object-related neurons.

Detailed scrutiny, focusing attention to particular locations or objects, unbinds illusory conjunctions of features and rebinds the features veridically to identify items actually within the scene. Thus, *vision with scrutiny* is required to unbind initial incorrect conjunctions and revise *vision at a glance* when unexpected conjunctions are present in the scene.

In summary, results concerning initial scene perception suggest that we perceive categorical information at a glance (using high-level cortical mechanisms) and miss (or assume) details that are detected by lower areas but not represented in individual high-level receptive

fields. Only later do we consciously perceive these details veridically by focusing serially on components and features, slowly scanning them one at a time.

## Visual Search
### Two Search Modes
Twenty years ago, Treisman noted that there are two modes of visual search, which she called "feature" or pop-out search versus "conjunction" or serial search (Treisman and Gelade, 1980; Treisman, 1988, 1993). As illustrated in Figures 2A and 2B, when the target element is defined by a single distinctive feature (such as a large difference in color, orientation, or size from distractor elements), its detection is rapid and parallel (independent of the number of distractors). On the other hand, when the target element differs only a little or only in a conjunction of features or parts, its detection is often set-size dependent, as if requiring serial focused attention, as demonstrated in Figures 2C–2F. According to Treisman's Feature Integration Theory (FIT), simple features are registered in parallel in specialized subsystems. Focused attention is needed to serially scan, integrate, and bind these features. When attention is overloaded, illusory conjunctions are formed, i.e., veridical binding of features requires focused attention. Thus, spread attention, which suffices for feature detection, contrasts with focused attention required for identifying conjunctions (Treisman and Gormican, 1988; Treisman, 1993, 1999). A basic tenet of FIT is that initial vision identifies elements without registering their precise location.

The properties of feature search, including its speed, automaticity, and locality, together with the elementary nature of the features that were found to pop-out, led to the suggestion that the odd element is detected at low-level processing stages where neuronal responses were believed to be largely determined by retinal input (Desimone and Ungerleider, 1989). The dependence of feature detection on local gradients seemed to imply that these stages may be as early as areas V1–V3 (Sagi and Julesz, 1987; Nothdurft, 1991; Nothdurft et al., 1999), and this notion was supported by physiological (Knierim and Van Essen, 1992; Kastner et al., 1997) and anatomical (Merigan et al., 1993) studies. However, associating initial feature search with low cortical areas has been questioned in a number of ways: low-level neuronal mechanisms have precise position information, while pop-out does not (Atkinson and Braddick, 1989); they discriminate fine orientation or color differences that do not pop-out, and their receptive fields are too small for lateral inhibition among distractors to be the source of pop-out. In addition, the basis for this association was the presumed "automaticity" of pop-out, while recent studies indicate that attention, albeit spread, is required for feature search, i.e., it depends on limited resources, since simultaneous performance of another task may interfere with it (Joseph et al., 1997; Ahissar et al., 2001). Furthermore, while some of the simple features that pop-out (such as orientation, color, motion, and depth) are represented in V1, others are not—including closure and geometric shape (e.g., circles versus rectangles; we relate to pop-out of more complex features below).

### Reverse Hierarchy of Visual Search
We propose the opposite conclusion, that feature search reflects high cortical level activity based on large, spread-attention receptive fields. This leads to position- and size-invariant feature detection, as suggested by the Feature Integration Theory (Treisman and Gormican, 1988). We suggest that pop-out is another manifestation of rapid but high-level *vision at a glance*. Note that the time required to enable observers to report an odd element is about the same as that needed to detect a complex category in an RSVP task, namely, about 100 ms (as measured by stimulus-to-mask onset asynchrony rather than reaction time; see Figure 3A). According to this view, the features that pop out and their characteristics will match those of high- but not low-level receptive fields. Indeed, over the years, anecdotal evidence came to light presenting exceptions to the rule that feature search should be possible only for basic features. Furthermore, pop-out clearly follows a variety of cortical computations.

Many high-level features pop out, as demonstrated in Figures 7 and 8. These include depth from shading (Figure 7D; Ramachandran, 1988), facial expressions (Ohman et al., 2001), and one's own face (Tong and Nakayama, 1999), 3D features (Figure 7B; Enns and Rensink, 1990), surface planes (He and Nakayama, 1992), parts and wholes (Wolfe et al., 1994), and perceptual groups (Bravo and Blake, 1990).

Feature search depends not on measurable spatial characteristics, but on feature categories, represented at high-level cortex. Even the simplest features that pop-out, such as orientation, incorporate category-like characteristics rather than direct metric properties (e.g., steep versus shallow or left versus right tilted). For example, oddly oriented bars that are mirror images of the distracting bars do not pop-out (e.g., it is harder to notice a bar oriented at $-50°$ from the vertical on a field of $+50°$ oriented distractors than to find one of $+10°$ on the same background; Wolfe et al., 1992). Similarly, the presence of a Vernier offset pops out of distractors with abutting lines, but a right-offset will not pop-out in a field of left-offset Verniers (Fahle, 1990). Treisman and others reported a large number of asymmetries in rapid visual search. Oblique bars pop-out from an array of vertical bars, orange from among reds, "C"s among "O"s, and so on, but not vice versa (Treisman and Gormican, 1988). These were understood as deriving from there being a limited set of canonical feature maps, representing primary orientations, colors, and special features such as circles and gaps. Wolfe et al. (1992) summarized many findings of this nature, concluding that features that pop-out are better described in terms of categories than as measurable spatial characteristics. We now propose that selection of canonical forms reflects high cortical level characteristics and are natural outcomes of the categorization of properties found there.

Feature search involves discriminating between the odd element and elements that are similar to their neighbors. As such, it must involve mechanisms that respond to many neighboring elements. However, it has been found that feature search is largely scale independent (Julesz, 1986). Thus, if one views an array at a given distance and an oddly oriented element pops-out, a similar pop-out will occur with viewing at a large range
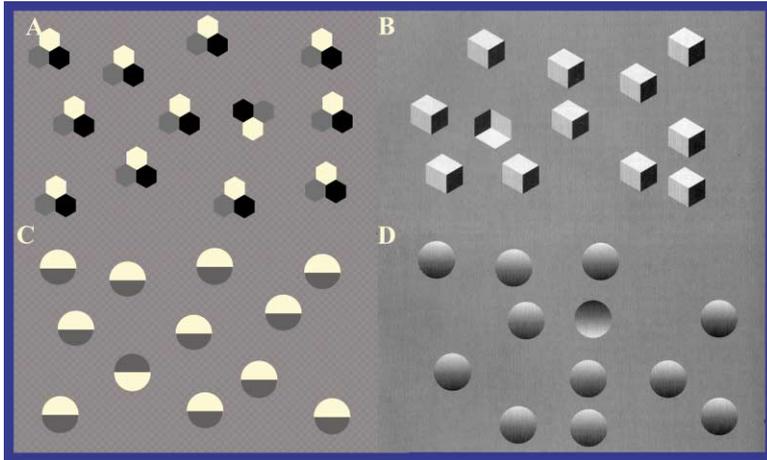
Figure 7. Evidence for High-Level Basis of Feature Search

Complex features also pop out, including 3D shape ([B], from Enns and Rensink, 1990) and depth from shading ([D], from Ramachandran, 1988) but not similar luminance arrangements (A and C).

of distances. Although the apparent size remains largely the same, retinal size of each element can vary, and pop-out remains. Furthermore, most pop-out presentations utilize fixed-size elements, although cortical size differs greatly as one scans different eccentricities, and still pop-out is obtained. In general, feature search works for a vast range of spatial scales, sizes, and interelement distances (once local grouping has been eliminated; Bacon and Egeth, 1991), including values that are far larger than predicted from the small receptive fields of low-level areas. All these conditions are consistent with a high-level rather than a low-level attribution.

Evidence for placing feature search higher rather than lower in the hierarchy comes also from findings that illusion effects act prior to pop-out mechanisms. Thus, pop-out is determined by elements' perceived or apparent size rather than by the size of their retinal representation. For example, the shape "+" pops-out of a field of "L"s if the bars of each are the same retinal length (Bergen and Julesz, 1983). However, the appearance of the bars are very different in the two cases, with those on the "+"s appearing considerably shorter due to the bisected line illusion. If one extends the "+" bars so that they *appear* as long as those of the "L"s, then the "+" no longer pops-out (Figure 8A; Bergen and Adelson, 1988). Similarly, salience of the odd element in feature search is affected by length, as determined by the Mueller-Lyer illusion (Figure 8B; Rensink and Enns, 1995), and by depth, as determined by texture gradient (Aks and Enns, 1996).

Other studies show that pop-out also follows completion processes. When a scene includes partially occluded shapes or objects, there is rapid completion of the fragments so that completed shapes are quickly recognized. Most striking, however, is the finding that the fragments themselves—though they are obviously the cues to the presence of the completed forms, do not pop-out, that is, they are not immediately available to conscious perception (Figure 8C; Rensink and Enns, 1995, 1998). Only with focused scrutiny can we become aware of these forms. We conclude, therefore, that low cortical mechanisms do not support explicit pop-out; only high cortical areas can underlie pop-out.

### Do Faces Pop-Out?
Assuming that feature search reflects higher-level receptive field properties, we would expect pop-out for categories of objects, which are represented at these levels. In an attempt to map the behavioral pop-out phenomenon to neuronal representations, we hypothesize that for a target object to pop-out it must be represented by a neuronal population that does not overlap with the population representing the distractors. The current understanding of high-level neuronal representation supports one such segregation clearly: recent fMRI studies indicate that face representation is largely separated from that of other objects, in particular, houses (Kanwisher et al., 1997; Levy et al., 2001). We would thus expect pictures of faces to pop-out on the background of houses.

To test this prediction, Elgavi-Hershler and Hochstein (O. Elgavi-Hershler and S. Hochstein, 2002, Vis. Sci. Soc., abstract) gathered drawings or photographs of houses, cars, and faces and measured the dependence of target detection on the number of distractor elements (Figure 9). Reaction time for detecting a picture with a face on a background of pictures of houses or cars (even when distractors differed in shape and size) was independent of set size, as expected from parallel feature search. Thus, this important high-level category does pop-out. In this demonstration, pop-out could be attributed to some basic feature that differentiates houses from faces (e.g., curvature). However, when trying to control for roundness and all other "special" features in faces (or houses), it was very difficult to retain the immediate perception of "faceness" (or "houseness"). Thus, simple, low-level detected features may indeed facilitate pop-out of high-level categories. We suggest that it is exactly the property of being relevant for high-level categorization that yields their salience in explicit perception. That is, we believe that it is the implicit use of basic features in explicit recognition of high-level categories that renders these features capable of producing pop-out: circles pop-out of squares because faces pop-out of houses, not vice versa. Figure 9 (bottom right) demonstrates that affecting the high-level interpretation of a picture (by scrambling elements of a schematic face so that it no longer looks like a face) induces pop-out, while totally erasing one element while leaving its face-like nature does not. In this way, Reverse Hierarchy Theory proposes an a priori concept regarding which features will pop-out, namely, those that underlie differences between basic categories. Indeed, we also
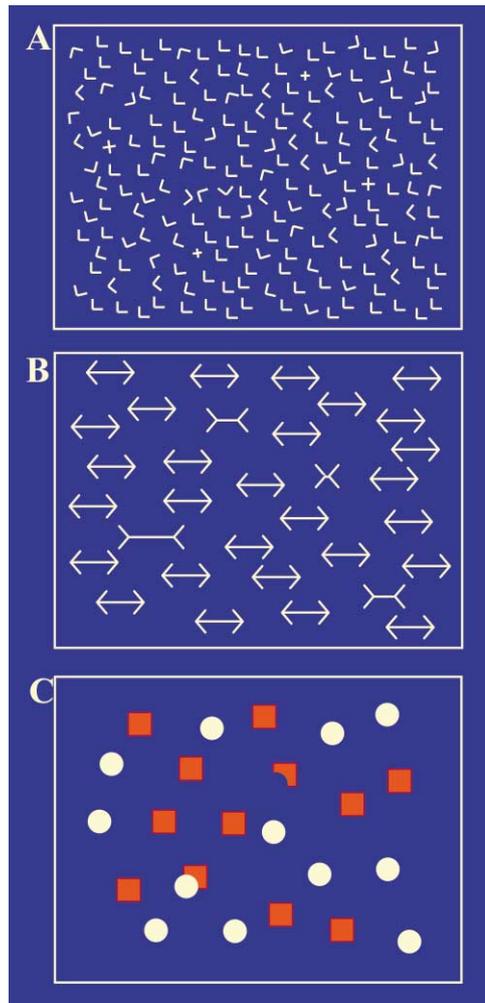
Figure 8. Evidence that Feature Search follows Considerable Implicit Processing

(A) Due to the crossed-line illusion, the lines of "+"s appear shorter than do those of "L"s. Thus, "+"s pop out among "L"s, except when their lines are stretched (from Bergen and Adelson, 1988).
(B) Perceived length—and pop-out—are determined by the Miller-Lyer illusion (Rensink and Enns, 1995).
(C) Due to amodal completion, the partially occluded square does not pop out, though a similarly shaped element does (Rensink and Enns, 1998).

found pop-out when using photographs rather than line drawings of faces, houses, and cars.

**Low-Level Cortical Involvement**
**in Conjunction Search**
Additional support for the new view comes from difficult conjunction search having the quantitative characteristics of low-level cortical areas. First and foremost, the small receptive fields of low-level neurons are natural candidates for serving as the substrate for *focused-attention vision with scrutiny*. Of course, it is conceivable that higher levels recompute fine position information. However, a more efficient method would be the use of feedback pathways to retrieve the information that is readily available there. Appropriate low-level locations may be selected by detecting active bottom-up inputs.

It has long been noted that search based on fine differences in basic dimensions is very similar to conjunction search in reaction time and set-size dependence. Thus, while a large difference in orientation or color pops-out, a small difference requires search with scrutiny. We propose that both conjunction search and fine detail discrimination search depend on low-level area neurons. For example, the minimal orientation or color difference that is sufficient for successful search with focused-attention scrutiny is far smaller than that which yields rapid parallel search (orientation detection with prolonged scrutiny: $<5°$, Vogels and Orban, 1985; orientation pop-out: $>20°$, Bergen and Julesz, 1983; Nothdurft, 1991; color, Nagy and Sanchez, 1991). Similarly, serial search is needed to find an element surrounded by distractors with orientations (or colors) that flank that of the target (Figure 10B, D'Zmura, 1991; Figure 10D, right, Wolfe et al., 1992). Thus, the receptive fields of V1 neurons support the values for scrutiny detection far better than do those of IT, and the opposite is the case for pop-out detection.

Even in the context of fast pop-out orientation detection, if observers are required to specify the exact target position, responses become closer to slow conjunction search. That is, observers need less time to identify target orientation than to accurately localize it (Atkinson and Braddick, 1989), supporting our conclusion that explicit perception gains access rather late to low-level computations where spatial localization information is encoded. Finally, evidence for low-level conjunction search comes from tests with a partially split-brain patient (where high-level area connections were spared). Distractors added to the hemifield without the target had no effect on conjunction search times, suggesting that search was accomplished in low-level, functionally split areas (H. Esteky et al., 2001, Soc. Neurosci., abstract).

**Physiological Evidence for Reverse Hierarchy**
Reverse Hierarchy Theory has direct predictions for the dynamics of activity in low-level cortical areas. Initial activity following stimulus initiation should be stimulus driven, localized, and automatic. It is expected to reflect the basic bottom-up implicit processing leading to the first *vision at a glance* percept in higher-level cortex. However, activity should continue with reentrant feedback along the reverse hierarchy, when and if *vision with scrutiny* is activated. This later activity would reflect top-down effects, including space- and object-based attention. These predictions are consistent with a range of recent findings in a variety of behavioral paradigms including feature search.

Reentry to lower areas may be investigated by unit recording in the behaving primate or by imaging techniques. Recent studies were successful in finding significant attention effects in lower areas, including V1 (Motter, 1993; Rosenthal and Hochstein, 1994; Gandhi et al., 1999; Sengpiel and Hubener, 1999) and early extrastriate areas (Treue and Maunsell, 1996; O'Craven et al., 1997; Hillyard et al., 1998). Thus, even primary visual cortex may participate in figure-ground segregation (Lamme, 1995), attentive and conscious scene analysis (Lamme and Spekreijse, 2000), context information (Zipser et al., 1996), and high-level vision (Lee et al., 1998) via feed-
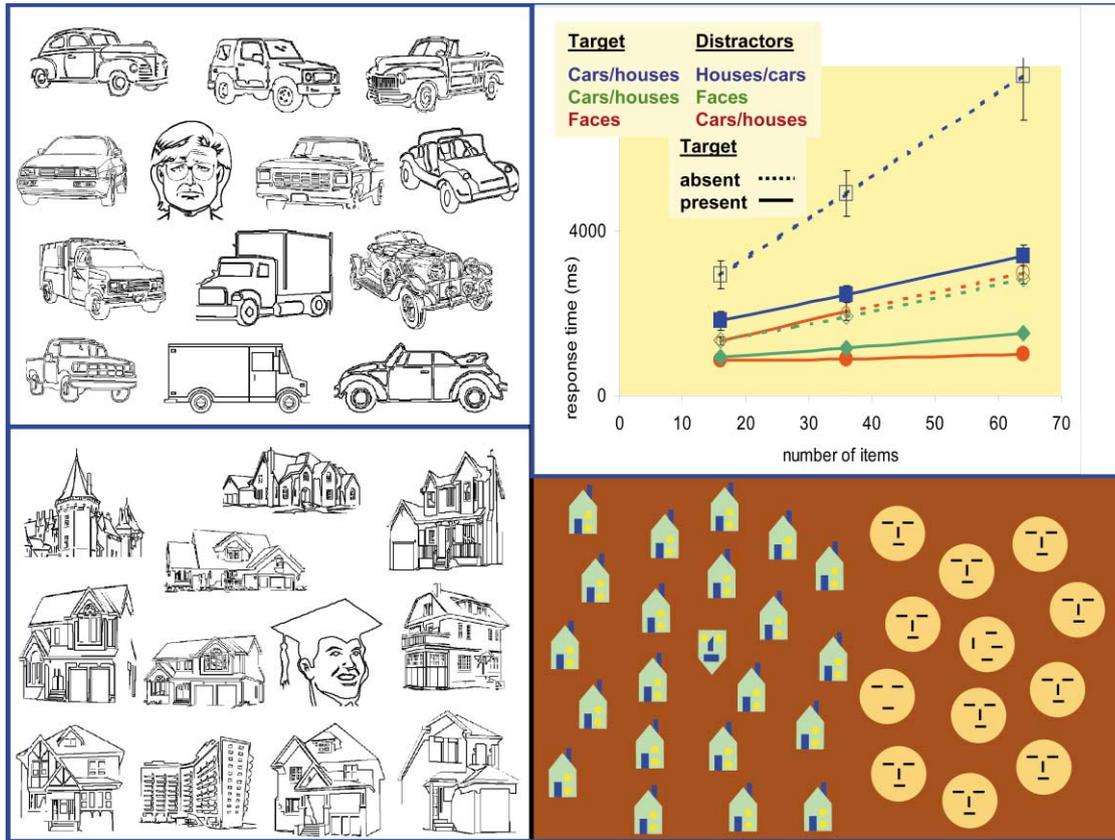
Figure 9. Search for Faces Is Easy as Predicted by Easy Search Being High Level

Subjects reported the presence or absence of an odd element in an array of 16, 36, or 64 elements presented on a computer monitor. Targets and distractors were line drawings of faces, houses, or cars, as demonstrated on the left for face targets among car or house distractors. The graph shows performance as a function of set size. Note that with faces as target or distractor, there is no dependence on set size for target presence trials (red and green solid lines) and a small slope for target absent trials (dashed lines)—the signature of easy feature search. For car or house target on house or car distractors, there are larger slopes in both cases (O. Elgavi-Hershler and S. Hochstein, 2002, Vis. Sci. Soc., abstract). Bottom right: with schematic elements, using rearranged parts to make up the face and house images, search is still easy. In addition, a non-face (with scrambled elements) pops-out, but a noseless face does not.
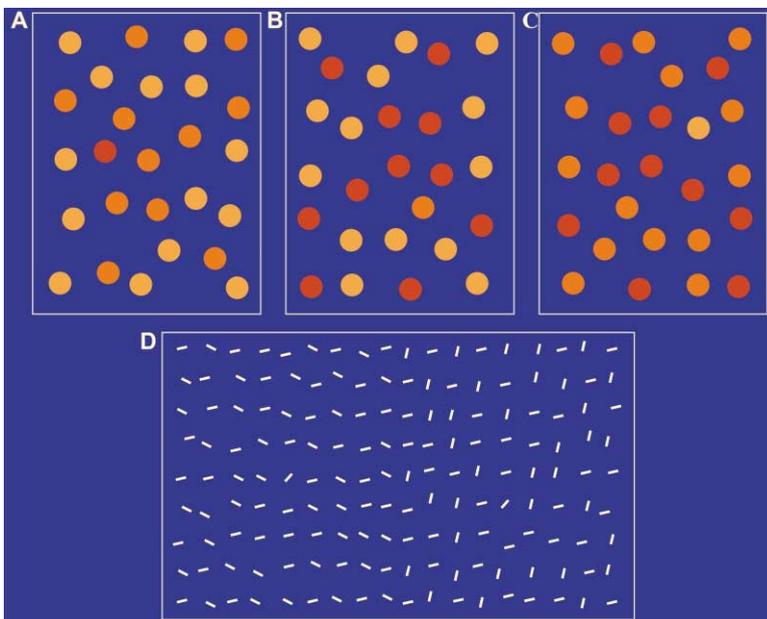


Figure 10. When Two Distractors Are Present, Search Is Hard when the Target Lies between Them in Feature Space but Is Easy Otherwise

(A–C) Red (or yellow) pops out of orange and yellow (or red) distractors, but orange does not pop out of red and yellow distractors (from D'Zmura, 1991).

(D) 45° pops out of 15° and 315° distractors but not from 15° and 75° distractors (from Wolfe et al., 1992).

**Table 1. Comparison of Old and New Theories of Conscious Visual Perception**

| Mechanism | Common View | | RHT Proposal | | Evidence |
|---|---|---|---|---|---|
| | Site | Time | Site | Time | |
| Gist of scene perception; basic level categorization | high | late | high | early | RSVP phenomena; rapid scene detection; illusory conjunctions |
| Subordinate categories | high | late | low | late | High areas represent basic categories; no evidence for still higher areas representing subordinate categories |
| Focused attention | high | late | low | late | Late attention effects in V1; "inattentional" blindness; object-centered attention → top-down guided |
| Feature search pop-out; illusory conjunctions | low | early | high | early | High-level features and large j.n.d.; parallel search (spread attention) → large receptive fields; dual-task interference → attention; no free-floating features |
| Search for conjunctions and difficult discriminations | high | late | low | late | Discrimination j.n.d. matches low-level fine resolution |
| Easy condition perceptual learning | high | late | high | early | Rapid generalized learning; "eureka" enabling phenomena |
| Hard condition perceptual learning | high | late | low | late | Slow specific learning that requires prior enabling |

back processing (Li et al., 2000; Pascual-Leone and Walsh, 2001). Newer studies failed to replicate V1 figure-ground dependence, though they did find long-latency responses to texture stimuli well outside the classical receptive field (Rossi et al., 2001). Most importantly, new reports suggest that attention affects only late portions of low area responses (Nobre et al., 1998; Roelfsema et al., 1998; Martinez et al., 1999) via feedback processing (Li et al., 2000; Pascual-Leone and Walsh, 2001). V1 attention effects are consistent not only with a space-based spotlight of attention but also with object-based attention (Somers et al., 1999; Tipper and Behrmann, 1996). The simple hierarchical model associating low-level cortex with automatic early vision and high-level cortex with later attentional perception does not fit these new findings. Instead, top-down attention effects at low-level areas seem to be subsequent to initial object perception in higher-level cortex, as predicted by Reverse Hierarchy Theory.

Recent physiological findings (Lee et al., 2002) directly support the high-level site of feature search and top-down reentrant aspect of later involvement of low-level areas. These investigators recorded single-unit activity in V1 and found that neurons responded equally to fields of elements whether or not they were of the type that produced behavioral pop-out. Displays with a single element produced a large response, but multi-element arrays with or without an odd element produced only a smaller response. However, after monkeys were trained to remember the precise location of the pop-out and use this information to direct a subsequent saccade, then the response to the pop-out type stimulus was enhanced to the level of that of the single element—though only belatedly (>120 ms after onset).

Thus, as predicted by Reverse Hierarchy Theory, V1 was *not* used for pop-out, but its activity reflected later availability of specific detailed information. Instead, the ability to detect orientations quickly must stem from the representation of orientation gradients at high-level areas. Such representations were found by single-unit methods (Vogels and Orban, 1996) and large-scale imaging techniques. In higher-level areas, they also sub-

serve the ability to define objects defined by texture gradients (Malach et al., 1995).

**Discussion**
In conclusion, we suggest a reversal of the way of understanding conscious perception and its relationship to cortical mechanisms. Based on results from feature search, *vision at a glance* and *vision with scrutiny* have been viewed as reflecting, respectively, low-level and high-level cortical representations. Thus, effortless simple feature detection has been seen as reflecting mechanisms operating at lower levels. Subsequent studies finding that the pop-out phenomenon also occurs for complex features challenged this view, while accumulating evidence for global precedence was viewed as an oddity.

We propose instead that *vision at a glance* reflects high-level mechanisms, while *vision with scrutiny* reflects a return to low-level representations. Reversing the anatomical associations puts many sets of seemingly conflicting data in a single coherent framework, as illustrated in Table 1. Thus, early spread attention reflects the large receptive fields found in high-level areas, and focused attention reflects localized low-level representations. High-level spread attention subserves our initial, crude global percept of the gist of the scene. Pop-out is but one aspect of this crude initial assessment. Associating early conscious perception with high cortical level mechanisms has implications for attentional phenomena as well. For example, Feature Integration Theory is an "early selection" theory (Broadbent, 1958), meaning that attention is required to bind elementary features into objects (Treisman, 1999). In the RHT view, early high-level perception makes a first approximation guess at binding, and focused attention—by return to low levels—is required to confirm or correct this guess and to veridically bind features. Since *vision at a glance* allows for spread attention parallel processing, RHT may resemble a "late selection" theory (Deutsch and Deutsch, 1963), e.g., allowing for attention to objects (Duncan, 1984; Vecera and Farah, 1994). However, RHT predicts that when attention is focused down to specific low-

level cortical activity, default high-level detection may be compromised, and parallel activity may go unnoticed (as in inattentional blindness; see above). Taken together, RHT predictions may be more in line with a controlled parallel processing model (Pashler, 1998).

The idea of late access to low-level information is part of several models incorporating the massive feedback pathways found anatomically. For example, Edelman (1989) suggested a theory of vision where reentry plays an important role. More recently, Ullman (1995) proposed that a possible function of cortical counterstreams might be a bidirectional seeking of sequences of activation. It was suggested that backward masking is due to reentry signals arriving at low levels after the input has changed (Di Lollo et al., 2000) and that versatile input filters are dynamically reconfigured by prefrontal cortex to handle incoming stimuli with maximum efficiency (Di Lollo et al., 2001). RHT is novel in its specifically relating implicit and explicit perception to mechanisms in the feedforward and reverse-hierarchy, feedback pathways. Incorporating the implicit/explicit and *vision at a glance/ vision with scrutiny* dichotomies into a specific anatomical framework of cortical hierarchies resolves many open perceptual puzzles regarding the nature of early and later perception.

This new view may be related to the long-term debate regarding direct versus indirect perception. von Helmholtz's (1867) theory of unconscious inference proposed that perception depends on the use of stored information to infer implications of an ambiguous retinal image, a view that received recent amendment and support (Gregory, 1970; Hochberg, 1981; Ullman, 1980; Rock, 1997). Theorists of the Gestalt (Koffka, 1935) and Gibsonian schools (Gibson, 1979) contested this view and claimed that perception is direct and not requiring inferential processing—including immediate perception of global complex features present in the external scene though not obvious in the retinal image. The source of the conflict may be that perception often *seems* immediate and direct, whereas theory and physiology support an indirect process. RHT proposes that initial *vision at a glance* includes results of automatic and implicit, bottom-up processing—making initial explicit perception introspectively direct, i.e., without conscious antecedents. Later *vision with scrutiny* may include more indirect conscious perceptual constructs. Whereas much information is available all along at lower cortical levels, focused attention is needed to bring it into consciousness.

In the search for neural correlates of consciousness (NCC), two models have been proposed (reviewed in Crick and Koch, 1998). First, consciousness may be correlated with activity in particular neuronal groups, perhaps at special sites. The thalamus, layers 5 and 6 of high-level cortical areas, and prefrontal cortex have each been suggested. Alternatively, consciousness may reflect a special type of activity in the very same neurons whose activity is not otherwise available to consciousness. Suggestions for the distinctiveness of the NCC include activity that is prolonged, correlated with activity in other regions, or phase locked to cortical oscillations. Reverse Hierarchy Theory proposes that low area (perhaps even V1) activity initially plays an implicit role in building complex receptive fields of higher areas—which are in turn correlates of conscious perception of

the *gist of the scene*. Later, these same low-level areas form the correlates of explicit perception of scene details. This dual role is probably most easily accommodated by a special activity rather than a special area model.

## References

Ahissar, M., and Hochstein, S. (1993). Attentional control of early perceptual learning. Proc. Natl. Acad. Sci. USA *90*, 5718–5722.

Ahissar, M., and Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. Nature *387*, 401–406.

Ahissar, M., and Hochstein, S. (2000). The spread of attention and learning in feature search: effects of target distribution and task difficulty. Vision Res. *40*, 1349–1364.

Ahissar, M., and Hochstein, S. (2002). The role of attention in learning simple visual tasks. In Perceptual Learning, M. Fahle and T. Poggio, eds. (Cambridge, MA: MIT Press), pp. 253–272.

Ahissar, M., Laiwand, R., and Hochstein, S. (2001). Attentional demands following perceptual skill training. Psychol. Sci. *12*, 56–62.

Aks, D.J., and Enns, J.T. (1996). Visual search for size is influenced by a background texture gradient. J. Exp. Psychol. Hum. Percept. Perform. *22*, 1467–1481.

Ariely, D. (2001). Seeing sets: representation by statistical properties. Psychol. Sci. *12*, 157–162.

Atkinson, J., and Braddick, O. (1989). Where and what in visual search. Perception *18*, 181–189.

Bacon, W.F., and Egeth, H.E. (1991). Local processes in preattentive feature detection. J. Exp. Psychol. Hum. Percept. Perform. *17*, 77–90.

Balz, G.W., and Hock, H.S. (1997). The effect of attentional spread on spatial resolution. Vision Res. *37*, 1499–1510.

Bergen, J.R., and Adelson, E.H. (1988). Early vision and texture perception. Nature *333*, 363–364.

Bergen, J.R., and Julesz, B. (1983). Rapid discrimination of visual patterns. IEEE SMC *13*, 857–863.

Biederman, I., Rabinowitz, J.C., Glass, A.L., and Stacy, E.W., Jr. (1974). On the information extracted from a glance at a scene. J. Exp. Psychol. *103*, 597–600.

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

Broadbent, D.E. (1958). Perception and Communication (London: Pergamon Press).

Bushnell, M.C., Goldberg, M.E., and Robinson, D.L. (1981). Behavioral enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. J. Neurophysiol. *46*, 755–772.

Cave, K.R., and Wolfe, J.M. (1990). Modeling the role of parallel processing in visual search. Cognit. Psychol. *22*, 225–271.

Crick, F., and Koch, C. (1995). Are we aware of neural activity in primary visual cortex? Nature *375*, 121–123.

Crick, F., and Koch, C. (1998). Consciousness and Neuroscience. Cereb. Cortex *8*, 97–107.

Cumming, B.G., and Parker, A.J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. Nature *389*, 280–283.

Desimone, R., and Ungerleider, L.G. (1989). Neural mechanisms of visual processing in monkey. In Handbook of Neuropsychology,

Volume 2, F. Boller and J. Grafman, eds. (Amsterdam: Elsevier Science), pp. 267–299

Deutsch, J.A., and Deutsch, D. (1963). Attention: Some theoretical considerations. Psychol. Rev. 70, 80–90.

Di Lollo, V., Enns, J.T., and Rensink, R. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. J. Exp. Psychol. Gen. 129, 481–507.

Di Lollo, V., Kawahara, J., Zuvic, S.M., and Visser, T.A.W. (2001). The preattentive emperor has no clothes: a dynamic redressing. J. Exp. Psychol. Gen. 130, 479–492.

Duncan, J. (1984). Selective attention and the organization of visual information. J. Exp. Psychol. Gen. 113, 501–517.

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

Edelman, G. (1989). The Remembered Present (New York: Basic Books).

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

Fahle, M. (1990). Is Vernier displacement a texton? Invest. Ophthalmol. Vis. Sci. 31, S105.

Fahle, M., and Poggio, T. (2002). Perceptual Learning (Cambridge, MA: MIT Press).

Fiorentini, A., and Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. Nature 287, 43–44.

Gandhi, S.P., Heeger, D.J., and Boynton, G.M. (1999). Spatial attention affects brain activity in human primary visual cortex. Proc. Natl. Acad. Sci. USA 96, 3314–3316.

Gibson, J.J. (1979). The Ecological Approach to Visual Perception (Boston, MA: Houghton Mifflin).

Goodale, M.A., and Milner, A.D. (1992). Separate visual pathways for perception and action. Trends Neurosci. 15, 20–25.

Gregory, R.L. (1970). The Intelligent Eye (London: Weidenfeld).

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

Hillyard, S.A., Vogel, E.K., and Luck, S.J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. Philos. Trans. R. Soc. Lond. B Biol. Sci. 353, 1257–1270.

Hochberg, J. (1981). On cognition in perception: perceptual coupling and unconscious inference. Cognition 10, 127–134.

Hock, H.S., Schoener, G., and Hochstein, S. (1996). Perceptual stability and the selective adaptation of perceived and unperceived motion directions. Vision Res. 36, 3311–3323.

Hock, H.S., Balz, G.W., and Smollon, W. (1998). Attentional control of spatial scale: effects on self-organized motion patterns. Vision Res. 38, 3743–3758.

Hubel, D.H., and Wiesel, T.N. (1968). Receptive fields and functional architecture of monkey striate cortex. J. Physiol. 195, 215–243.

Humphreys, G.W. (1998). Neural representation of objects in space: a dual coding account. Philos. Trans. R. Soc. Lond. B Biol. Sci. 353, 1341–1351.

Intraub, H. (1999). Understanding and remembering briefly glimpsed pictures: Implications for visual scanning and memory. In Fleeting Memories, V. Coltheart, ed. (Cambridge, MA: MIT Press), pp. 47–70.

Johnston, J.C., and McClelland, J.L. (1974). Perception of letters in words: Seek not and ye shall find. Science 184, 1192–1194.

Joseph, J.S., Chun, M.M., and Nakayama, K. (1997). Attentional requirements in a 'preattentive' feature search task. Nature 387, 805–807.

Julesz, B. (1986). Texton gradients: the texton theory revisited. Biol. Cybern. 54, 245–251.

Kanwisher, N.G. (1987). Repetition blindness: type recognition without token individuation. Cognition 27, 117–143.

Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311.

Kastner, S., Nothdurft, H.C., and Pigarev, I.N. (1997). Neuronal correlates of pop-out in cat striate cortex. Vision Res. 37, 371–376.

Knierim, J.J., and Van Essen, D.C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. J. Neurophysiol. 67, 961–980.

Koffka, K. (1935). Principles of Gestalt Psychology (New York: Harcourt Brace).

Kooi, F.L., De Valois, K.K., Switkes, E., and Grosof, D.H. (1992). Higher-order factors influencing the perception of sliding and coherence of a plaid. Perception 21, 583–598.

Lamme, V.A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. J. Neurosci. 15, 1605–1615.

Lamme, V.A., and Spekreijse, H. (2000). Modulations of primary visual cortex activity representing attentive and conscious scene perception. Front. Biosci. 5, D232–D243.

Lee, T.S., Mumford, D., Romero, R., and Lamme, V.A. (1998). The role of the primary visual cortex in higher level vision. Vision Res. 38, 2429–2454.

Lee, T.S., Yang, C.F., Romero, R.D., and Mumford, D. (2002). Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. Nat. Neurosci. 5, 589–597.

Levy, I., Hasson, U., Avidan, G., Hendler, T., and Malach, R. (2001). Center-periphery organization of human object areas. Nat. Neurosci. 4, 533–539.

Li, W., Their, P., and Wehrhahn, C. (2000). Contextual influence of orientation discrimination of humans and responses of neurons in V1 of alert monkeys. J. Neurophysiol. 83, 941–954.

Liu, Z. (1999). Perceptual learning in motion discrimination that generalizes across motion directions. Proc. Natl. Acad. Sci. USA 96, 14085–14087.

Livingstone, M., and Hubel, D. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. Science 240, 740–749.

Lumer, E.D., and Rees, G. (1999). Covariation of activity in visual and prefrontal cortex associated with subjective visual perception. Proc. Natl. Acad. Sci. USA 96, 1669–1673.

Luo, C.R., and Caramazza, A. (1996). Temporal and spatial repetition blindness: effects of presentation mode and repetition lag on the perception of repeated items. J. Exp. Psychol. Hum. Percept. Perform. 22, 95–113.

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

Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R., and Tootell, R.B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc. Natl. Acad. Sci. USA 92, 8135–8139.

Martinez, A., Anllo-Vento, L., Sereno, M.I., Frank, L.R., Buxton, R.B., Dubowitz, D.J., Wong, E.C., Hinrichs, H., Heinze, H.J., and Hillyard, S.A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. Nat. Neurosci. 2, 364–369.

Maunsell, J.H., and Newsome, W.T. (1987). Visual processing in monkey extrastriate cortex. Annu. Rev. Neurosci. 10, 363–401.

Merigan, W.H., Nealy, T.A., and Maunsell, J.H.R. (1993). Visual effects of lesions of cortical area V2 in macaques. J. Neurosci. 13, 3180–3191.

Moran, J., and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. Science 229, 782–784.

Most, S.B., Simons, D.J., Scholl, B.J., Jimenez, R., Clifford, E., and Chabris, C.F. (2001). How not to be seen: The contribution of similarity and selective ignoring to sustained inattentional blindness. Psychol. Sci. 12, 9–17.

Motter, B.C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. J. Neurophysiol. 70, 909–919.

Nagy, A.L., and Sanchez, R.R. (1991). Critical color differences determined with a visual search task. J. Opt. Soc. Am. A 7, 1209–1217.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. Cogn. Psychol. 9, 353–383.

Nobre, A.C., Allison, T., and McCarthy, G. (1998). Modulation of

human extrastriate visual processing by selective attention to colours and words. Brain *121*, 1357–1368.

Nothdurft, H.C. (1991). Texture segmentation and pop-out from orientation contrast. Vision Res. *31*, 1073–1078.

Nothdurft, H.C., Gallant, J.L., and Van Essen, D.C. (1999). Response modulation by texture surround in primate area V1: correlates of "pop out" under anesthesia. Vis. Neurosci. *16*, 15–34.

O'Craven, K.M., Rosen, B.R., Kwong, K.K., Treisman, A., and Savoy, R.L. (1997). Voluntary attention modulates fMRI activity in human MT-MST. Neuron *18*, 591–598.

Ohman, A., Lundqvist, D., and Esteves, F. (2001). The face in the crowd revisited: a threat advantage with schematic stimuli. J. Pers. Soc. Psychol. *80*, 381–396.

Pascual-Leone, A., and Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. Science *292*, 510–512.

Pashler, H.E. (1998). The Psychology of Attention (Cambridge, MA: MIT Press).

Potter, M.C. (1976). Short-term conceptual memory for pictures. J. Exp. Psychol. [Hum. Learn.] *2*, 509–522.

Ramachandran, V.S. (1988). Perception of shape from shading. Nature *331*, 163–166.

Ramachandran, V.S., and Anstis, S.M. (1983). Perceptual organization in moving patterns. Nature *304*, 529–531.

Rensink, R.A., and Enns, J.T. (1995). Preemption effects in visual search: evidence for low-level grouping. Psychol. Rev. *102*, 101–130.

Rensink, R.A., and Enns, J.T. (1998). Early completion of occluded objects. Vision Res. *38*, 2489–2505.

Rensink, R.A., O'Regan, J.K., and Clark, J.J. (1997). To see or not to see: the need for attention to perceive changes in scenes. Psychol. Sci. *8*, 368–373.

Rock, I. (1997). Indirect Perception (Cambridge, MA: MIT Press).

Roelfsema, P.R., Lamme, V.A., and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. Nature *395*, 376–381.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., and Boyes-Bream, P. (1976). Basic objects in natural categories. Cognit. Psychol. *8*, 382–439.

Rosenthal, O., and Hochstein, S. (1994). Effects of stimulus meaning on V1 orientation tuning properties. Israel Soc. Neurosci. *3*, 47.

Rossi, A.F., Desimone, R., and Ungerleider, L.G. (2001). Contextual modulation in primary visual cortex. J. Neurosci. *21*, 1698–1709.

Rubin, N., Nakayama, K., and Shapley, R. (1997). Abrupt learning and retinal size specificity in illusory-contour perception. Curr. Biol. *7*, 461–467.

Sagi, D., and Julesz, B. (1987). Short-range limitation on detection of feature differences. Spat. Vis. *2*, 39–49.

Sengpiel, F., and Hubener, M. (1999). Visual attention: spotlight on the primary visual cortex. Curr. Biol. *9*, R318–R321.

Somers, D.C., Dale, A.M., Seiffert, A.E., and Tootell, R.B. (1999). Functional MRI reveals spatially specific attentional modulation in human visual cortex. Proc. Natl. Acad. Sci. USA *96*, 1663–1668.

Squire, L.R. (1986). Mechanisms of memory. Science *232*, 1612–1619.

Tipper, S.P., and Behrmann, M. (1996). Object-centered not scene-based visual neglect. J. Exp. Psychol. Hum. Percept. Perform. *22*, 1261–1278.

Tong, F., and Nakayama, K. (1999). Robust representations for faces: Evidence from visual search. J. Exp. Psychol. Hum. Percept. Perform. *25*, 1016–1035.

Treisman, A. (1986). Features and objects in visual processing. Sci. Am. *255*, 106–115.

Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. Q. J. Exp. Psychol. *40A*, 201–237.

Treisman, A. (1993). The perception of features and objects. In Attention: Selection, Awareness and Control, A. Baddeley and L. Weiskrantz, eds. (Oxfork, UK: Oxford University Press), pp. 5–35.

Treisman, A. (1999). Solutions to the binding problem: progress through controversy and convergence. Neuron *24*, 105–125.

Treisman, A., and Gelade, G. (1980). A feature integration theory of attention. Cogn. Psychol. *12*, 97–136.

Treisman, A., and Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. Psychol. Rev. *95*, 15–48.

Treue, S., and Maunsell, J.H. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. Nature *382*, 539–541.

Ullman, S. (1980). Against direct perception. Behav. Brain Sci. *3*, 373–415.

Ullman, S. (1995). Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. Cereb. Cortex *5*, 1–11.

Ungerleider, L.G., and Mishkin, M. (1982). Two cortical visual systems. In Analysis of Visual Behavior, D.J. Ingle, M.A. Goodale, and R.J.W. Mansfield, eds. (Cambridge, MA: MIT Press), pp. 549–580.

Vecera, S., and Farah, M.J. (1994). Does visual attention select objects or locations? J. Exp. Psychol. Gen. *123*, 146–160.

Vogels, R., and Orban, G.A. (1985). The effect of practice on the oblique effect in line orientation judgments. Vision Res. *25*, 1679–1687.

Vogels, R., and Orban, G.A. (1996). Coding of stimulus invariances by inferior temporal neurons. Prog. Brain Res. *112*, 195–211.

von Helmholtz, H. (1867). Treatise on Physiological Optics, Volume 3, J.P.C. Southall, transl. (1962) (New York: Dover).

Wolfe, J.M., Friedman-Hill, S.R., Stewart, M.I., and O'Connell, K.M. (1992). The role of categorization in visual search for orientation. J. Exp. Psychol. Hum. Percept. Perform. *18*, 34–49.

Wolfe, J.M., Friedman-Hill, S.R., and Bilsky, A.B. (1994). Parallel processing of part-whole information in visual search tasks. Percept. Psychophys. *55*, 537–550.

Wong, E., and Weisstein, N. (1982). A new perceptual context-superiority effect: line segments are more visible against a figure than against a ground. Science *218*, 587–589.

Zipser, K., Lamme, V.A., and Schiller, P.H. (1996). Contextual modulation in primary visual cortex. J. Neurosci. *16*, 7376–7389.