
The Modeling and Control of Visual Perception

Ronald A. Rensink

When we view our surroundings, we invariably have the impression of experiencing it via a “picture” formed immediately and containing a great amount of detail. This impression is the basis of three strong intuitions about how visual perception works: (1) Because visual experience is immediate, it must result from a relatively simple system. (2) Because the picture we experience is unitary, perception must involve a single system whose only goal is to generate this picture. (3) Because this picture contains enough information to let us react almost immediately to any sudden event in front of us, it must contain a complete description of almost everything in sight.

But recent research has shown that each of these intuitions is wrong: (1) The immediacy of an output is no guarantee that the underlying processes are simple. In fact, recent work has shown visual perception to be a highly complex activity, with a considerable amount of sophisticated processing done extremely rapidly. (2) If we experience a unitary percept, this does not necessarily mean that a single integrated system created it.

Indeed, visual perception increasingly appears to involve several quasi-independent subsystems, only some of which are responsible for the picture we experience. (3) If we can quickly access information about something whenever needed, this does not imply that all of it is represented at all times. Indeed, recent work shows that the picture we experience contains far less information at any given moment than our conscious impressions indicate, particularly in regard to dynamic events.

This new view of vision, then, suggests that the processes involved are more sophisticated than previously believed, and may do more than just provide a picture to our minds. The view provides considerable support for dynamic models that emphasize interaction and coordination of component processes and that have a strong sensitivity to what the operator knows and the task they are engaged in. It also points toward the possibility that vision itself might be controlled in interesting ways, allowing its component processes to be seamlessly incorporated into systems that extend beyond the physical body of the operator.

This chapter provides an overview of this new view of vision, focusing on several of the results and theories that have emerged. The first section, “Component Systems,” discusses individual processes, characterizing them in terms of how they relate to attention. The second proposes how these internal systems might be integrated to produce the picture we consciously experience. The section “Integration of External Systems” then discusses some possible ways that these integration mechanisms might interact with external systems, creating more effective forms of human–machine interaction.

Component Systems

Visual perception results from the operation of a highly complex and heterogeneous set of processes, some of which remain poorly understood to this day. This section briefly describes several of these processes, along with some of the theories and models put forward to account for their operation. (For a more complete discussion, see Palmer, 1999.)

Since many of the new findings about visual perception involve attention—either in terms of what it is or how various operations relate to it—processes are

grouped here into three largely disjoint sets: those that act before visual attention operates, those involved with attention itself, and those in which attention—and perhaps consciousness—may never be involved at all.

Preattentive Processes

When light enters the eye, it strikes the retina and is transformed into an array of neural signals that travels along the optic nerve, maintaining a retinotopic organization. What happens next is less well understood, but the prevailing view is that this marks the beginning of *early vision* (Marr, 1982), a stage of vision characterized by processes that are *low level* (i.e., operating locally on each point of the retinotopic input) and *rapid* (i.e., completed within about 200 ms). These are believed to operate automatically, without any need for attention (Figure 10.1).

Simple Properties

Early visual processing is thought to create a set of “primitives” on which all subsequent processing is based. Information concerning the nature of these primitives has largely been obtained via two kinds of study. The first is texture perception (e.g., Julesz, 1984), where

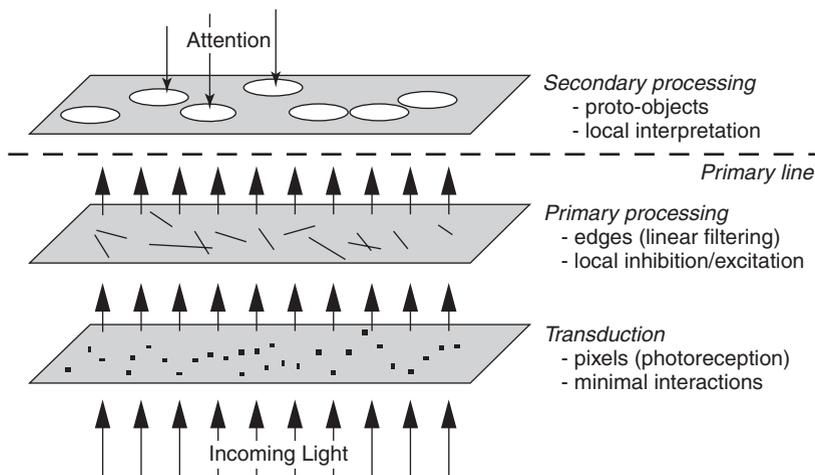


FIGURE 10.1 Schematic of early visual processing. The first stage is *transduction*, where photoreception occurs (i.e., the retina). The next is *primary processing*, where linear or quasi-linear filters measure image properties. This is followed by *secondary processing*, which applies “intelligent” nonlinear operations. Processing in all three stages is carried out rapidly and in parallel across the visual field. The outputs of the secondary stage constitute the 2½D sketch. The contents of this sketch are also the operands for subsequent attentional processes—the limit to immediate attentional access is given by the *primary line* (see Rensink, 2000).

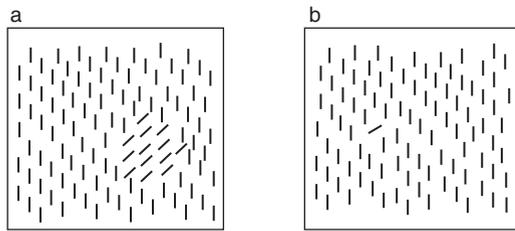


FIGURE 10.2 Tests for visual primitives. (a) Texture perception. Elements differing in orientation lead to the effortless segmentation of patch from the rest of the display. (b) Visual search. An item with a unique orientation immediately “pops out” from among the other items.

textons are defined as elements whose properties support effortless texture segmentation (Figure 10.2a). The second is visual search (e.g., Treisman, 1988; Wolfe, Cave, & Franzel, 1989; chapter 8, this volume), where basic *features* are properties that “pop-out”; that is, they can be quickly detected if they have a value unique to the display (Figure 10.2b).

In both cases, the set of primitives includes color, motion, contrast, and orientation. These properties have a common computational nature in that they can be determined on the basis of the limited local information around each point; this allows them to be computed rapidly and in parallel across the image (see chapters 8 and 9, this volume). The explicit redescription of the image in terms of such elements is sometimes referred to as a *primal sketch* (Marr, 1982). Most models of the underlying processes involve an initial stage of linear filtering, followed by various nonlinear operations (see Palmer, 1999).

Complex Properties

Although many early visual properties are simple, the structures they describe may be relatively complex. For example, an isolated line fragment will pop-out if it has a distinctive length. But what structure is measured to get this value? If a distinctive line segment becomes part of a group (e.g., a drawing), pop-out is no longer guaranteed—this will now depend on the overall length of the group (Rensink & Enns, 1995). This indicates two things. First, some forms of grouping occur preattentively, with “length” being that of the overall group. As such, this indicates a fair degree of *visual intelligence* at this level. Second, the components of a group are inaccessible to higher-level processes, at least over the periods of time characteristic of this stage. As such,

the elements of early vision may be better characterized as *proto-objects* (i.e., precursors of objects) rather than the outputs of the very first stages of processing (Rensink, 2000). Features would then be those properties of proto-objects that are capable of affecting performance.

Another example of such visual intelligence is the ability to compensate for occlusion. For example, if a bar is occluded by a cube midway along its length, the visible portions will project to two line segments separated by the image of the occluder. But these segments can be linked preattentively, the resultant proto-object reflecting that they correspond to the same object in the scene (Rensink & Enns, 1998). Scene-based properties themselves can also be encoded. For example, search can be affected by three-dimensional orientation, direction of lighting, and shadow formation, with these estimates apparently formed on the basis of “quick and dirty” assumptions true only most of the time (Rensink & Cavanagh, 2004). Such behavior is in accord with theories that postulate the goal of early vision to be a $2\frac{1}{2}D$ *sketch*, a viewer-centered description of the world in which scene properties are represented in a fragmented way (Marr, 1982).

Control

Most models of early vision assume a unidirectional flow of information from the retina to the $2\frac{1}{2}D$ sketch, without any influence from higher-level factors, such as the nature of the task or knowledge about the scene (Marr, 1982). But anatomical studies show that there are a huge number of return connections from higher levels to lower ones, and psychological studies have shown these connections to have perceptual consequences, resulting in the image of an item being “knocked out” of iconic memory under the right conditions (DiLollo, Enns, & Rensink, 2000). As such, the possibility arises that the set of elements at this stage may not be invariant for all tasks, but might be at least partly subject to higher-level control.

This possibility receives some support on computational grounds, in that it would be needlessly complex to have dedicated early-level processes for every possible aspect of scene structure. Rather, it might be more efficient to simply invoke instructions to calculate these whenever necessary.

Other Open Issues

Commonality of Visual Elements. The properties that govern texture segmentation are neither a subset

nor a superset of the properties that govern pop-out in visual search (Wolfe, 1992). This is difficult to reconcile with a single set of basic elements. It may be that different systems are involved; each with its own set of elements.

Reference Frame. Most models of early vision assume it to be based on a retinotopic frame of reference (see Palmer, 1999). However, visual search appears unaffected by sudden changes in position or size of the display, suggesting that it may be based upon a more abstract spatiotopic frame that is invariant to changes in size, and perhaps to other transformations as well (Rensink, 2004a).

Influence of Top-Down Control. Few experiments to date have investigated how the operation of early vision might be affected by the knowledge of the observer or the task they are carrying out. It has been proposed that early vision is not susceptible to these factors (Pylyshyn, 2003). But it is not clear how this can be reconciled with results showing the effects of connections from higher levels (DiLollo et al., 2000). Perhaps only particular types of control are possible.

Attentional Processes

Although observers can easily understand a request to “pay attention,” it has proved extraordinarily difficult to determine what is happening when they do so. Earlier models considered attention to be a unitary faculty, or homogeneous “stuff.” However, an emerging view is that *visual attention* is better characterized as the selective control of information in the visual system, which can be carried out in various ways by various processes (Rensink, 2003). As such, there exist—at least functionally—several kinds of attention, which may or may not be directly related to each other. (For an extensive set of current perspectives on attention, see Itti, Rees, & Tsotsos, 2005).

Selective Access

The simplest form of attention is *selective access*—the selective routing of some aspect of the input (usually involving a simple property) to later processes. For example, observers can detect a target more quickly and more accurately if cued to its location (Posner, Snyder, & Davidson, 1980). This has been explained by a *spotlight of attention* that amplifies inputs from the selected area and/or suppresses inputs from others.

It was once thought that selective access simply protected processors at higher levels from being overwhelmed by the sheer amount of information at early levels. More recently, selective access has been thought to serve a number of additional purposes. For example, it can improve the signal-to-noise ratio of the incoming signal and so improve performance (Treisman & Gormican, 1988). It can also delimit control of various actions—for example, by focusing on the particular part of an item to be grasped (Neumann, 1990). Appropriate use of selective access can also enable low-complexity approximations of high-complexity visual tasks (Tsotsos, 1990).

Selective Integration

Another form of attention is *selective integration*—the binding of selected parts or properties into more complex structures. For example, experiments have shown that it is difficult to detect a single L-shaped item among a set of T-shaped items. Similar difficulties are experienced for unique combinations of orientation and color, or of most other features. An influential account of this is *feature integration theory* (Treisman, 1988), which posits that attention acts via a spotlight that integrates the features at each location into an *object file*. If an item contains a unique feature, this spotlight is drawn to it automatically and the item is seen; otherwise, the spotlight must travel from item to item, integrating the corresponding feature clusters at a rate of about 50 ms/item.

The earlier belief was that to make good use of limited processing “resources” such integration needed to be selective. However, more recent work tends to view integration in terms of the *coordination* of the processes involved with each feature; selection can greatly simplify the management of this (Rensink, 2003).

Selective Hold Change Detection

Recent work shows that observers can have difficulty noticing a sudden change made simultaneously with an eye movement, a brief flash in the image, or a sudden occlusion of the changed item (e.g., Figure 10.3).¹ Such *change blindness* (Rensink, O’Regan, & Clark, 1997) occurs under a variety of conditions and can occur even when the changes are large, repeatedly made, and in full knowledge that they will occur. It can be accounted for by the hypothesis that attention is necessary to see change. A change will then be difficult to see whenever



FIGURE 10.3 Flicker paradigm. One way to induce change blindness is by alternating an original and a modified version of an image, with a brief blank or mask between each presentation. Performance is measured by time required to see the change. Even at a rate of two to three alternations per second, observers typically need several seconds to see the change. (In this example, the appearance/disappearance of the aircraft engine.)

the motion transients that accompany it cannot draw attention to its location (e.g., if they are swamped by other motion signals in the image).

In this view, seeing a change requires a selective process that involves several steps: (1) the item from the original image is entered into a short-term store—presumably visual short-term memory; (2) it is held there briefly; and (3) it is then compared with the corresponding item in the new image. Selection could potentially occur at any or all of these stages. It is currently unclear which—if any—is the critical one.

One proposed account is *coherence theory* (Rensink, 2000), in which attention corresponds to the establishment of coherent feedback between lower-level proto-objects and a higher-level collection point, or *nexus* (see “Integration of Component Systems” section). Here, attention is characterized by the *selective hold* of items in a short-term store. In addition to change detection (essentially the tracking of an item across time), this form of attention may also be involved in the tracking of items across space (Pylyshyn, 2003). The reason for its selectivity may arise from the difficulty of establishing or maintaining two or more distinct feedback circuits

that connect arbitrary (and possibly disparate) parts of the brain.

Control

Visual attention is subject to two different kinds of control. The first is *exogenous* (or low level), which automatically draws attention to a particular item or location. Exogenous control is governed by *saliency*, a scalar quantity that reflects the priority of attentional allocation. Saliency is usually modeled as a function of the spatial gradient of early-level features (e.g., changes in the average orientation in a certain part of the field), with large changes in density (such as those at the borders of objects or regions) having the highest saliency (Itti, 2005). Exogenous control is believed to be largely independent of high-level factors. Some aspects may be affected by task and instruction set, although this has not yet been firmly established (Egeth & Yantis, 1997; Theeuwes & Godijn, 2002).

The second kind of control is *endogenous* (or high level). This is a slower, more effortful form of control that is engaged voluntarily on the basis of more abstract,

context-sensitive factors such as task instruction. The relation of both exogenous and endogenous control to the various types of attention has not been worked out completely, nor has the way that exogenous and endogenous control interact (see Egeth & Yantis, 1997).

Other Open Issues

Relation Between Attention and Eye Movements. Experiments show that attention does not have to coincide with eye fixation: People can shift their attention without moving their eyes. Some form of attention is needed to select the target of an eye movement, but this need not be accompanied by a withdrawal of attention from other items. Thus, attention (of any type) and eye fixation are distinct processes and should not be conflated. However, a detailed model of the interaction between attention shifts and eye movements has not yet been developed (see Henderson, 1996).

The Basis of Selection. Visual search can be guided via the selection of features such as color or motion (Wolfe, Cave, & Franzel, 1989). For location, the situation is less clear: selection could be of a particular point in space (at which some object happens to be), an object (at some point in space), or perhaps both. Both spatial and object factors appear to play a role (see Egeth & Yantis, 1997). The relation of space- and object-based selection is not yet clear; these may be related to different types of attention.

Capacity. For space-based selection, a natural model is the spotlight of attention. Some models allow the intensity and size of the spotlight to be varied continuously; others do not (see Cave & Bichot, 1999). Most models posit one spotlight, although recent experiments suggest that it can be divided for some tasks (McMains & Somers, 2004). For object-based processes, capacity is usually about four items (Pylyshyn, 2003), although for some operations it is only one (Rensink, 2001, 2002a). There is currently no general consensus as to how all these accounts can be reconciled.

Nonattentional Processes

In the past, it was generally believed that the sole purpose of vision was to produce a sensory experience of some kind (i.e., a picture) and that attention was the “central gateway” to this. However, evidence is increasing that a good deal of sophisticated processing can be

done without attention even beyond the early stage and that some of these processes can result in outputs having nothing to do with visual experience.

Rapid Vision

Although recent work shows a considerable amount of visual intelligence at early levels (see “Preattentive Processes” section), this is not limited to processes based on local information. Instead, such intelligence is found throughout *rapid vision*—that aspect of perception carried out during the first few hundred milliseconds throughout the entire visual system, as the initial sweep of information travels from the eyes to the highest cortical levels, and perhaps back down again (Rensink & Enns, 1998).²

One quantity determined this way is the abstract meaning of a scene, or *gist* (e.g., whether it is a city, port, or farm). Gist can be ascertained within 100 ms of presentation, a time insufficient for attending to more than a few items. It can be extracted from blurred images and without attention; indeed, two different gists can be determined simultaneously. Gist is likely determined on the basis of simple measures such as the distribution of line orientations or colors in the image; other properties of early-level proto-objects may also be used. Some aspects of scene composition, such as how open or crowded it is, can also be obtained this way, without the involvement of coherent object representations. (For a more comprehensive review, see Oliva, 2005.)

A possibly related development is the finding that observers are extremely good at obtaining *statistical summaries* of a group of briefly presented items. For example, observers can match the average size of a group of disks to an individual disk about as accurately as they can match the sizes of two individual disks (Ariely, 2001).

“Medium-Term” Memory

Although attention (in the form of selective hold) appears to be involved with visual short-term memory, there also appears to exist a form of memory that does not require attention, at least for its maintenance. This is not the same as long-term memory, since it can dissipate after a few minutes, once there is no further need for it. As such, it might be considered to be a separate, *medium-term* memory.

One possibility in this regard is memory for *layout*—the spatial arrangement of objects in the scene, without regard to their visual properties or semantic identities

(Hochberg, 1968). Some layout information may be extracted within several seconds of viewing—likely via eye movements—and can be held over intervals of several seconds without a constant application of attention. (Tatler, 2002). Interestingly, memory for repeated layouts can be formed in the complete absence of awareness that such patterns are being repeated (Chun & Jiang, 1998); such memory appears to help guide attention to important locations in that layout.

Visuomotor Guidance

It has been proposed (Milner & Goodale, 1995) that vision involves two largely separate systems: a fast *on-line* stream, concerned with the guidance of visually guided actions such as reaching and eye movement, and a slower *off-line* stream, concerned with the conscious perception and recognition of objects. Evidence for this *two-systems theory* is largely based on patients with brain damage: some can see objects but have great difficulty grasping them, while others cannot see objects, but (when asked to) can nevertheless grasp them easily and accurately.

Effects also show up in normal observers. For example, if a dot is shifted the moment an observer moves his eye to it, his eye always makes a corrective jump to the new location, even if the observer has no awareness of the shift (Bridgeman, Hendry, & Stark, 1975). And if a target is displaced during an eye movement, the hand of an observer reaching toward it will correct its trajectory, even if the observer does not consciously notice the displacement (Goodale, Pelisson, & Prablanc, 1986).

Implicit Perception

Although somewhat controversial, consensus is increasing that *implicit* processes exist; that is, performance is affected in some way, even though no conscious picture of the stimuli is involved. An example of this is *subliminal* perception, where a stimulus is embedded among irrelevant items and presented very briefly (typically less than 50 ms), making it difficult to see. Presentation of an unseen stimulus under these conditions nevertheless has several effects, such as speeding the conscious recognition of it in a subsequent display. (For a discussion of this, see Norretranders, 1999.) It is thought that subliminal perception may be a form of perception without attention (Merikle & Joordens, 1997).

Results from other approaches are consistent with this position. In *inattention blindness*, observers can fail to see an unexpected stimulus if their attention is

focused elsewhere (Mack & Rock, 1998). Stimuli with strong emotional impact are exceptions, indicating that some degree of semantic processing can occur in the absence of attention.

Implicit perception can be explored by comparing performance for stimuli reported as “seen” against performance for stimuli reported as “unseen.” If these are not the same (e.g., have different sensitivities to color), the implicit processes must differ from those that gave rise to the conscious picture (Merikle & Daneman, 1998). This approach has uncovered several distinctive characteristics of implicit processing, such as a strong response to emotionally charged stimuli, a sensitivity to semantic meaning (but not to geometric structure), and an inability to exclude stimuli, for example, an inability to choose any word other than the one subliminally presented (Merikle & Daneman, 1998).

Control

In principle, it might be possible to control various aspects of a nonattentive process. For example, if it is possible to control the kind of outputs at the preattentive level (see “Preattentive Processes”), it might also be possible to control the outputs of other rapid processes, such as statistical summaries of different kinds. Inputs to nonattentive processes might also be selectable, for example, visuomotor operations might be able to act only on stimuli of a particular color or shape. If attention is defined as a selective process, then it might be that various forms of nonconscious attention exist. But separating out these from the known forms of *conscious* attention would be difficult.

Other Open Issues

Commonality of Processes. Although the perceptual processes in this subsection have been grouped together on the basis of not involving attention, such a negative definition says little about how—or even whether—they are related to one another. Issues such as the extent to which these processes are based on similar elements or involve common reference frames are still to be investigated.

Mindsight. Some observers can have a “feeling” that a change is occurring, even though they do not have a visual picture of the change itself (Rensink, 2004b). Although this phenomenon—*mindsight*—has been replicated, disagreement exists about how to best interpret it

(Simons, Nevarez, & Boot, 2005). The mechanisms involved are poorly understood. One possibility is that mindsight is a form of alert, relying on nonattentional processes such as layout perception (Rensink, 2004b).

Summary

Several major advances have recently occurred in our understanding of individual visual processes. One of these is the finding that considerable visual intelligence exists at early levels, with sophisticated processing carried out rapidly in the absence of attention. This opens up the possibility of interfaces that allow tasks traditionally done by *high-level* thinking to be off-loaded onto these faster, less effortful, and possibly more capable systems (Card, Mackinlay, & Shneiderman, 1999; Ware, 2004). The findings that intelligence also exists in rapid vision and visuomotor guidance opens up even more possibilities (see “Integration of External Systems” section).

Another set of advances concerns the nature of attention itself. Attention seems neither as pervasive, powerful, or unitary as originally believed. However, it still appears to be critical for particular operations, such as integrating information from selected items. Importantly, a delineation of the various processes (or at least functions) grouped under this label is now emerging, with some understanding of the characteristics of each. Among other things, this provides a much better grounding for the design of interface systems in which attention must be used in an appropriate way (see “Integration of External Systems” section).

Interestingly, recent work also shows that vision may involve more than just attention and consciousness—that systems may exist that operate entirely without the involvement of either. The existence of such systems has major implications for the modeling and control of visual perception, in that they indicate that the conscious picture experienced by an observer is only a part of a wider-ranging system. To appreciate what this might mean, we must consider how these component systems might be integrated.

Integration of Component Systems

How can the component systems of vision be integrated such that an observer can experience a unitary picture of their surroundings? It was originally believed that these processes—acting via attention—constructed a complete, detailed description of the scene, for example,

accumulating representations in a visual buffer of high-information density (see Rensink, 2002a). Models of this kind, however, have great difficulty explaining induced failures of perception, such as inattentional blindness and change blindness (see “Component Systems” section).

Recent work tends to view the integration of systems in dynamic rather than static considerations—as *coordination* rather than *construction*. Among other things, this has the consequence that different people can literally see the same scene in different ways, depending on their expectations and the task they are engaged in.

Coherence Theory

To illustrate the idea of coordination, consider how it might apply to visual attention (or at least, to selective hold). According to some models, attention *welds* visual features into relatively long-lasting representations. But if so, why aren't all visible items welded within the first few seconds of viewing, allowing detection of all objects and events under all conditions?

Rather than assuming that focused attention acts by forming new structures that last indefinitely, it may be that it simply endows existing structures with a degree of coherence, with this lasting only as long as attention is directed at them. Developing this line of thought leads to a *coherence theory* of attention (Rensink, 2000).

Basics

Coherence theory is based on three related hypotheses (Figure 10.4):

1. Before attention, early-level proto-objects are continually formed rapidly and in parallel across the visual field. Although these can contain detailed descriptions and be quite complex (see “Component Systems” section), they are *volatile*, lasting only a few hundred milliseconds. As such, they are constantly in flux, with any proto-object simply being *replaced* when a new stimulus appears at its location.
2. Attention selects a small number of proto-objects from this flux and stabilizes them into an object representation. This is done via reciprocal links between the selected items and a higher-level *nexus*. The resulting circuit (or *coherence field*) forms a representation coherent across space and time. A new stimulus at the attended location is

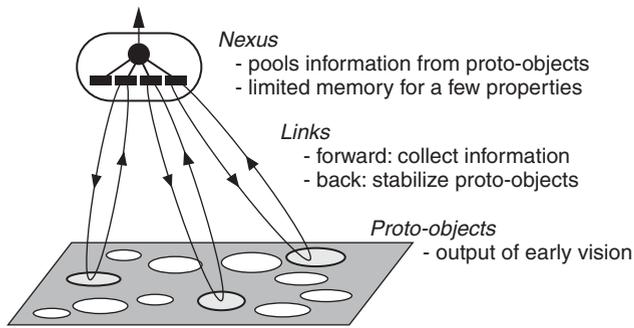


FIGURE 10.4 Coherence theory. Early vision continually creates proto-objects rapidly and in parallel across the visual field. Attention selects a subset of these, incorporating them into a circuit called the *coherence field*. As long as the proto-objects are “held” in this field, they provide the visual content of an individuated object that has both temporal and spatial coherence.

then perceived as the *change* of an existing structure rather than the appearance of a new one.

3. After attention is released (i.e., after the circuit is broken), the field loses its coherence and the object representation dissolves back into its constituent proto-objects. There is little or no *after-effect* of having attended to an item, at least in terms of the structures that underlie “here-and-now” perception.³ (Also see Wolfe, 1999.)

According to this view, then, attention is not stuff that helps create a static representation. Rather, it is the establishment (and maintenance) of a coordinated information flow that can span several levels of processing. The components that enter into the coherence field remain at the processing level where they were formed; what is added are the links that allow these components to be treated as part of the same object.

Implications

No Buildup of Attended Items. An important part of coherence theory is its assertion that attention corresponds to the establishment of a circuit of information flow. Thus, no buildup results from attention having been allocated to an item—once attention is withdrawn, the components of the coherence field revert to their original status as volatile proto-objects. Since attention is limited to just a few items (see Rensink, 2002a), most parts of a scene will therefore not have a coherent, detailed representation at any given moment. Of course, more durable, longer-term representations can exist, but these do not have such coherence.

No Complete Coherent Representation. According to coherence theory, some representations (those of early vision) are *complete*, in that they cover the entire visual field, and some representations (those resulting from

attention) are *coherent* over space and time. However since there is no buildup of coherence fields, no representation can be both complete *and* coherent.⁴

No Dense Coherent Representation. Although a proto-object might have a high density of information, only a small amount of this is held in the nexus (Rensink, 2001, 2002a). Thus, a coherence field cannot hold much detail about the attended item. If one of the properties represented in the nexus is one of the properties changing in the world, the change will be seen. Otherwise, it will not, even if the object is attended.

Virtual Representation

If only a few objects in a scene can have a coherent representation at any time, and if only a few properties of these objects are encoded, why do observers have the impression of seeing all events in the scene in great detail?

One way to account for this is the idea of a *virtual representation*: instead of a coherent representation of all the objects in an observer’s surroundings, a coherent representation is created only of the object—and its properties—needed for the task at hand (Rensink, 2000).

Basics

If a coherent representation of an object can be created whenever needed, and if this representation contains those aspects required for the task at hand, the representation of the scene will appear to higher levels as if *real*, as if all objects are represented in complete detail simultaneously. Such a representation will have all the power of a real one, while using much less in the way of resources.

This strategy has been successfully applied to information systems. For example, it is the basis of virtual

memory in computers, where—if coordination of memory access is successful—more memory appears available than is physically present at any given time. Browsing Web sites on a computer network can also be characterized this way (Rensink, 2000).

Requirements for Successful Operation

Virtual representation reduces complexity in space by trading it off for increased complexity in time. Only certain types of task can take advantage of this trade-off. For visual perception, what is required is

1. only a few objects need to be represented at any time, and
2. only a few properties of these objects need to be represented at that moment, and
3. the appropriate object(s) can always be selected, and
4. the appropriate information about that object is always available when requested.

The first requirement is easily met for most tasks. Most operators need to control only one object (e.g., a steering wheel) or monitor one information source (e.g., a computer display) at a time. Tasks involving several independent objects or events can usually be handled by time-sharing, that is, rapidly switching between the objects or events. The second requirement is likewise easily met, in that most tasks only involve a few properties of an object at any given time (e.g., its overall size or color). Time-sharing can again be used if several properties are needed. The third requirement can also be met, provided that three conditions hold. The first is having the ability to respond to any sudden event, and create the appropriate representation. As discussed in the section “Attentional Processes,” this ability—in the form of the exogenous control of attention—does exist in humans. The second is having the ability to anticipate events so that nothing important is missed, even if other events are occurring. This can be done if the observer has a good understanding of the scene (i.e., knows what to expect) to direct endogenous attentional control appropriately. Third, the average time between important events must be at least as great as the average switching time of the control mechanisms. This is generally true for the world in which we live (or, at least, our ancestral environment), where important events almost never occur several times a second.

The fourth requirement is also met under most conditions of normal viewing. Provided that eye fixation and attention can be directed to the location of a selected object and that sudden occlusions are not common, it will usually be possible to obtain visual detail from the stream of incoming light, with the relevant properties then extracted from this stream. Thus, a high-capacity internal memory for objects is not needed: detailed information is almost always available from the world itself, which acts as an *external repository* (or *external memory*).⁵

Implications

Dependence on Knowledge and Task. In this view, the visual perception of a scene—including all events taking place in it—rests on a dynamic “just in time” system that represents only what is needed for the task at hand. The degree to which this is successful will depend on how well the creation of appropriate object representations is managed. Since this in turn strongly depends on the knowledge of the observer and the task being carried out, different people will literally see the same scene in different ways.

Change Blindness Blindness. If the creation of object representations is managed well, virtual representation will capture most of the relevant events in an environment. Meanwhile, any failure to attend to an appropriate object will not be noticed, and will usually have few consequences. As such, observers will generally become susceptible to *change blindness blindness*—they greatly overestimate their ability to notice any large changes that might occur (Levin, 2002).

Distributed Perception. Virtual representation implies a partnership between observer and environment: rather than an internal re-presentation containing all details of the scene, the observer uses an external repository (i.e., the world), trusting that it can provide detailed information whenever needed. In an important sense, then, observer and surroundings form a single system, with perception *distributed* over the components involved.

Triadic Architecture

The successful use of virtual representation requires that eye movements and attentional shifts can be made to the appropriate object at the appropriate time. But how might this be done? One possibility consistent

with what is known of human vision is the *triadic architecture* (Rensink, 2000).

Basics

This architecture involves three separate systems, each of which operates somewhat independently of the others (Figure 10.5):

1. An *early vision* system that rapidly creates detailed, volatile proto-objects in parallel across the visual field (see “Preattentive Processes”).
2. A limited-capacity *attentional* system that links these structures into coherent object representations (see “Coherence Theory” section).
3. A nonattentional *setting system* that provides a context to guide attention to the appropriate objects (see “Nonattentional Processes” section).

These largely correspond to the groups of systems in the “Component Systems” section, except that the setting system contains only those nonattentional processes that control visual attention. In addition to these three systems, there is a connection to long-term knowledge—such as schemas and particular skills—that helps direct high-level (endogenous) control, and

so influences perception. But since most of long-term memory is effectively off-line at any instant it is not considered part of here-and-now visual perception (Rensink, 2000).

Of the three systems involved in this architecture, the one concerned with setting is perhaps the least articulated. It likely involves at least two aspects of scene structure useful for the effective endogenous control of attention:

1. The abstract meaning (or *gist*) of the scene, for example, whether it is a forest or barnyard (see “Nonattentional Processes” section). This quantity is invariant over different eye positions and viewpoints, and, to some degree, over changes in the composition of objects. Consequently, it could provide a stable constraint on the kinds of objects expected and perhaps even indicate their importance for the task at hand.
2. The spatial arrangement (or *layout*) of objects in the scene. This quantity is, at least from an allocentric point of view, invariant to changes in eye position, and as such could help direct eye movements and attentional shifts. If held in a medium-term memory, the location of many objects could be represented. Some additional information

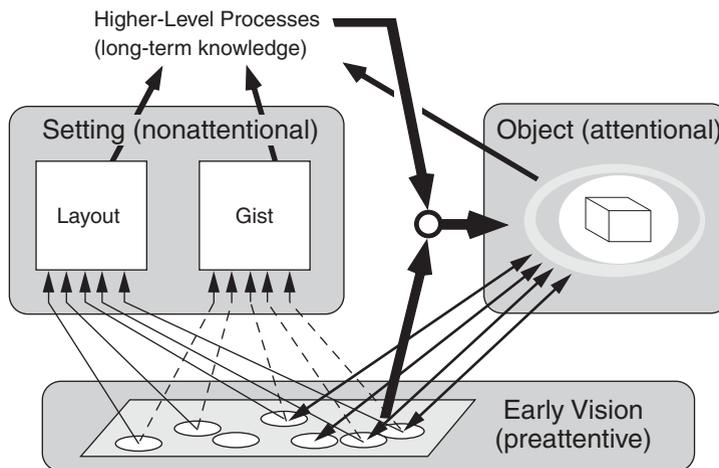


FIGURE 10.5 Triadic architecture. Visual perception is carried out by three interacting systems: (1) Early vision creates volatile proto-objects. (2) Attention “grabs” these structures and forms an object with temporal and spatial coherence. (3) Setting information—together with long-term knowledge and salience estimates obtained from early visual processing—guides attentional management.

concerning each item may also be possible; such information need not be extensive to be useful.

Interaction of Systems

Although there is currently little empirical evidence concerning the way that these systems interact, one possibility is as follows:

1. Early vision provides a constantly regenerating sketch of the scene visible to the observer.
2. Gist, layout, and perhaps some object semantics are determined without attention; these invoke a scene schema in long-term memory which provides constraints on the types of objects that might be present, possible actions, and so on.
3. The invoked schema is verified, beginning with a simple checking of expected features. Items consistent with the schema at this level need not be examined (and therefore need not be encoded) in detail.
4. If an unexpected structure in the image is encountered or an (unknown) salient item is suddenly detected at early levels, attentional processes form a coherent representation of it, attempt to determine its identity, and possibly reevaluate the gist. Layout can be used both to check the current interpretation as well as help guide attention to a requested object.

Such interaction involves a complex combination of exogenous and endogenous control, as well as of immediate, relatively changeable information about the scene and longer-term, more stable knowledge.

Implications

Construction Versus Coordination. Representations beyond early vision are no longer dense structures *constructed* via eye movements and attentional shifts; instead, they may be better viewed as sparse structures that *coordinate* the use of detailed information from the world. Thus, early-level representations are not *replaced* by more complex representations, but are *incorporated* into circuits spanning several levels of processing.

Role of Attentional Processing. Rather than being the *central gateway* of visual perception, attention (in the

form of selective hold) may only be one of several concurrent streams—the one concerned with the conscious perception of coherent objects. Other streams may operate in complete independence of it. Indeed, the role of attention (or at least of consciousness) itself may even be somewhat restricted in regards to the control of action, being mostly involved with initiation of actions in unfamiliar situations, and perhaps learning (see Norretranders, 1999).

Role of Nonattentional Processes. In this view, nonattentional processes do not rely on attention for their “intelligence”—they in fact help guide it. Nonattentional processes beyond the setting system may enable aspects of perception having nothing to do with the production of a conscious picture, such as affecting emotional state, or guiding visuomotor actions (see the “Integration of External Systems” section).

Summary

The integration of the component systems underlying visual perception appears to be achieved on the basis of *coordination* rather than *construction*. The nature of this more dynamic view can be seen, for example, in the coherence theory of visual attention. Here, attention is treated as a linkage between selected components, without the need for a separate representation of the attended object. The components are simply incorporated into a circuit along which information circulates, with a relatively small nexus serving to stabilize this linkage.

This style may also apply to our experience of a scene. Rather than being based on a static, dense representation, our experience may be based on a *virtual representation* that encodes only what is needed at each moment. As such, what is perceived will depend strongly on the particular observer and on the task they are carrying out. This account also leads to a view in which observer and environment form a single system, with perception distributed over the components involved.

One possible implementation of this in human vision is the *triadic architecture*, where a critical role is played by the mechanisms that control attention. Here, perception is distributed over several component systems, with the nonattentional (and presumably nonconscious) mechanisms that control attention playing a critical role. If these mechanisms can be properly controlled, it might be possible to integrate the component systems of human vision not only with each other but also with

external systems. The next section explores some of these possibilities.

Integration of External Systems

It increasingly appears that visual perception may be based on several quasi-independent systems, each with its own kind of intelligence (see “Component Systems” and “Integration of Component Systems” sections). Given that these systems are integrated via dynamic coordination rather than static fusion, and that this coordination can be influenced by external factors, the possibility arises that these systems can be integrated with external systems as well.

If so, consideration of the mechanisms that carry out this coordination would provide a basis for the design of more effective visual display systems. It also opens up some genuinely new prospects for human-machine interaction.

3.1 Reduced Change Blindness

Given that attention is needed to see change (see “Attentional Processes” section), an observer will be blind to most unattended transitions in a display, resulting in informative transitions being missed. This would be especially important in displays in which information is conveyed dynamically, for example, if an operator is tracking the location of an item or following the orientation of an indicator needle. During such times, a transition could easily occur elsewhere in the display, for example, an alert appearing, without the operator noticing it. High-level (endogenous) control could lower the likelihood of such change blindness, but even if the observer could maintain a full state of alertness, the likelihood of missing something will still be considerable.

Change blindness can be induced in a number of ways (see Rensink, 2002a); a system should reduce each of these contributions as much as possible. For example, change blindness can be induced by eye movements, which make up about 10% of total viewing time on average⁶; any transition will therefore have as much as a 10% chance of being missed because of this factor alone. One way of lowering this likelihood is by minimizing the need for (or the size of) eye movements, for example, by keeping important sources of information close together. In addition, displays could minimize the number of dynamic events occurring

elsewhere, since these could draw attention to themselves, diverting attention away from the main information source. Moreover, although a single event can be attended without problems, two cannot—their contents will be pooled (Rensink, 2002a). Consequently, only a single source of dynamic information should ever be used at any time.

Coercive Graphics

A more speculative possibility involves the use of non-conscious processing to control what the observer consciously experiences. Given that the visual experience of an observer depends on the coordination of attention, and given that this coordination is strongly affected by what is shown to the eyes, the possibility arises of *coercive graphics*—displays that can control attention to make the observer see (or not see) a particular part of the image (Rensink, 2002b).

Coercion has long been used by magicians and filmmakers to achieve a variety of striking effects. Three means of control are commonly used:

1. High-level interest. Semantic factors that influence the semi-voluntary control of attention, for example, stories that interest the observer in a particular object or event.
2. Mid-level directives. Cues that require some intelligence, but then cause attention to rapidly move to a given location. Examples are the direction of eye gaze of another person (or image), and the direction of finger pointing.
3. Low-level salience. Simple scalar quantity that is the basis of exogenous control (see “Attentional Processes” section). Attention is automatically drawn—often involuntarily—to items such as those with a unique color, motion, orientation, or contrast.

All of these can be highly effective when done by humans (see Sharpe, 1988). If a system could make effective use of these, it could lead to *magical displays* capable of effects even more powerful than those produced by professional magicians.

A coercive display could ensure that important events would not be missed. It might also speed up operation by directing attention to required locations or items. Coercion would also be useful for older observers, acting as a form of “glasses” to compensate

for the reduction in attentional abilities that generally happens with increasing age. Again, the user would notice nothing unusual—they would simply never miss anything important that occurred.

Emotional Control/Vigilance

In the past, visual displays were concerned only with the visual experience of the observer. But according to the triadic architecture (see “Triadic Architecture” section), this experience involves just one perceptual stream—the attentional system. However, other systems may also operate in tandem with this, and carry out a significant (albeit nonconscious) part of perception. As such, the potential arises for displays expressly designed to work with such processes, and influence aspects of an observer other than the visual percept they experience.

One such example is the control of emotional state. Nonattentional (and nonconscious) processes have a pronounced sensitivity to emotionally laden words and pictures (see “Nonattentional Processes” section). Moreover, some of these processes can affect the physiological mechanisms underlying the associated emotions, even though the stimuli involved are unseen (e.g., Liddell et al., 2005; Whalen et al., 1998). As such, it may be possible to develop displays that could, e.g., calm an operator down or increase their level of vigilance, all on the basis of stimuli that are not consciously experienced.

Soft Alerts

For many tasks, the system must allow the operator to respond quickly to unexpected events. This is typically done via an alert, which attempts to draw attention to its location, thereby ensuring it will be seen. Although such alerts can be successful, they can also be dangerous (especially for time-critical tasks), since they have the potential to divert attention away from important objects or operations.

A somewhat speculative alternative to these involves the phenomenon of *mindsight*—the feeling of something happening without an accompanying picture (see “Nonattentional Processes” section). This phenomenon is poorly understood; it may be related to feelings generated by emotional states, although this is far from certain (Rensink, 2004b). In any event, if visual displays could be designed to invoke this feeling whenever desired, it would make an extremely useful form of alert, a *soft alert*, that would not disturb existing attentional

control (Rensink, 2002b). Such an alert would be useful for situations where the arrival of a new event does not require immediate attention, for example, the arrival of email while the operator is monitoring a changing situation.

Direct Support of Action

As discussed in the “Nonattentional Processes” section, considerable evidence exists that actions such as reaching and grasping are guided by nonattentional systems having nothing to do with conscious visual experience (see Milner & Goodale, 1995). It is likely that activities such as moving a mouse or pointing are guided similarly. More generally, the set of visuomotor systems (along with other motor systems and perhaps some rapid perceptual processes) may be coordinated to result in an *inner zombie* capable of carrying out operations in a highly sophisticated way, even though consciousness is not involved (see Norretranders, 1999).

If this view is correct, it suggests the need for displays designed expressly for the *direct support* of action, that is, displays to act directly on the nonconscious visuomotor systems rather than only on the systems that produce conscious visual experience. For example, pointing without visual feedback may help a user aim a laser pointer at a given location, even though this is counterintuitive from the viewpoint of conscious perception (Po, Fisher, & Booth, 2003). In such situations, there may be no awareness that the display is providing such guidance; the user simply does the right thing.

Cognitive Extension

The type of dynamic representation discussed here is a special case of the more general notion of *deictic* (or *indexical*) representation. Here, the goal is not to *construct a copy* of the world, but rather to *coordinate* component systems so as to carry out actions in it (Ballard, Hayhoe, Pook, & Rao, 1997; Clancey, 1997; chapter 21, this volume). It does not matter whether the components involved are internal or external—all that matters is that they are part of a circuit of information flow under the control of the user (see Clark, 2003).

If the coordination with an operator’s visual system is done properly, external processors (e.g., a calculator or an information visualization system) could become part of such a circuit, allowing sophisticated processing to be incorporated in a seamless manner. External effectors

(e.g., a car or an airplane) could likewise become part of such a circuit, with each system treated as a visuomotor system of the operator. Indeed, when such an interface functions well, the operator can experience a literal extension of themselves into the task domain (e.g., becoming part of the car or airplane), resulting in highly effective control of all component systems (see Clark, 2003).

Summary

An emerging view is that the operation of the human visual system is based on the coordination of several quasi-independent systems. If the coordination mechanisms within an operator can be applied to external systems as well, highly effective forms of human-machine interaction could result. For example, systems might be designed to reduce the likelihood of change blindness, to ensure that the operator will always see what they need to see (assuming it is in the display), or even to bring other internal systems (e.g., emotions) into play. In addition, the possibility also exists of incorporating external systems—not only information sources but also processing elements and effectors—in a similar fashion, allowing human perceptual and cognitive abilities to be extended in a highly natural way.

Conclusions

This chapter has surveyed some of the main developments that have recently occurred in our understanding of human perception and discussed some of their implications for the modeling and control of visual perception. Among these developments is the increasing recognition that visual attention may not be the central gateway to visual perception but may instead be simply one of several quasi-independent systems, each capable of sophisticated processing. It also appears that these systems are not integrated via dense, static representations that accumulate results but rather via dynamic coordination that depends on such factors as the knowledge of the observer and the nature of the task they are engaged in. This kind of coordination—if done properly—can result in a virtual representation that provides the observer with a unitary picture of the scene. Such coordination may also enable purely non-conscious systems to act coherently, resulting in an *inner zombie* capable of intelligent on-line control of actions without any involvement of conscious.

From this viewpoint, then, human visual perception appears to be based on the coordination of several quasi-independent systems, each with its own form of intelligence. It may be useful to view human-machine interaction in a similar way, with the operation of a human-machine system based on the coordination of several quasi-independent systems (some internal to the operator, some external), each with *its* own form of intelligence. Such a perspective not only suggests ways of improving existing display and control systems, but also points to new possibilities for increasing the effectiveness and scope of human-machine interaction.

Acknowledgments

I would like to thank Nissan Motor Co. (Japan) and Natural Sciences and Engineering Research Council (Canada) for supporting the work described in this chapter. Thanks also to Wayne Gray, Chris Myers, and Hans Neth for their helpful comments on an earlier version. I would also like to thank Wayne Gray and Air Force Office of Scientific Research (USA) for giving me the opportunity to present these ideas in this context.

Much of the work described here—both the particular results and the general approach—has its origins during the years 1994–2000 at Cambridge Basic Research (CBR), a laboratory of Nissan Motor Co. in Cambridge Massachusetts. Thanks to my colleagues at CBR for their encouragement and support during that time: Jack Beusmans, Erwin Boer, Jim Clark, Rob Gray, Andy Liu, Simon Rushton, and Ian Thornton. Also thanks to Takao Noda and Akio Kinoshita for maintaining a wonderful environment that greatly fostered research. This chapter is dedicated to them.

Notes

1. This and other examples can be downloaded from www.cs.ubc.ca/~rensink/flicker/download or www.psych.ubc.ca/~rensink/flicker/download.

2. *Rapid vision* can be defined as that occurring during the first 200 ms or so of visual processing; it can involve mechanisms throughout the visual system. *Low-level vision* occurs via low-levels mechanisms, which generally operate in parallel in a spatiotopic array and without any influence of stimulus-specific knowledge. *Early vision* can be defined as the intersection of these two, that is, processing that is both rapid and low level (Rensink & Enns, 1998). As such, rapid

vision comprises several different—and coordinated—processing systems, of which early vision is one. *Preattentive processes* are the set of processes at early levels; these operate without attention, before any attentional application. Although all preattentive processes are nonattentional, not all nonattentional processes are preattentive.

3. There may be effects such as entry into long-term memory. But long-term memory is not considered to be among the mechanisms that directly underlie “here-and-now” (or “working”) perception (Rensink, 2000).

4. It may be that a relatively complete representation of the static aspects of a scene is built up—experimental evidence to date is not sufficient to rule out this possibility (Simons & Rensink, 2005). However, the existence of change blindness clearly shows the existence of severe limits on how much of its *dynamic* aspects are represented at any time.

5. The more usual term is *external memory* (see, e.g., Clark, 2003). But *memory* does not entirely capture the situation because what is available from the world is not a remnant of any information that disappeared from the environment. Even if information might have disappeared from the *observer*, it is still problematic how *remnant* would apply before or during the first time the observer accessed this information.

6. The duration of each ballistic movement (or *saccade*) of the eye depends on the angle A traversed, according to $D = 21 + 2.2A$, where D is in milliseconds, and A is in degrees (Carpenter, 1988). Such movements can sometimes take more than 100 ms. However, on average these take about 30 ms, at an average rate of about three to four per second (see Palmer, 1999). Thus, the amount of time spent in ballistic movement—where blur induced by the eye movement destroys the automatic drawing of attention to the location of a change—is typically 90–120 ms per second, with greater durations for movements through greater angles.

References

- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*, 157–162.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*, 723–767.
- Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, *15*, 719–722.
- Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). Information visualization. In S. K. Card, J. D. Mackinlay, & B. Shneiderman B. (Eds.), *Readings in information visualization: Using vision to think* (Chap. 1). San Francisco: Morgan Kaufman.
- Carpenter, R. H. S. (1988). *Movements of the eyes* (2nd ed., p. 72). London: Pion.
- Cave, K. R., & Bichot, N. P. (1999). Visuo-spatial attention: Beyond a spotlight model. *Psychonomic Bulletin and Review*, *6*, 204–223.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28–71.
- Clancey, W. J. (1997). *Situated cognition: On human knowledge and computer representations*. Cambridge: Cambridge University Press.
- Clark, A. J. (2003). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*. Cambridge, MA: MIT Press.
- DiLollo, 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.
- Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, *48*, 269–297.
- Goodale, M. A., Pelisson, D., & Prablanc, C. (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature*, *320*, 748–750.
- Henderson, J. M. (1996). Visual attention and the attention-action interface. In K. Akins (Ed.), *Perception* (pp. 290–316). Oxford: Oxford University Press.
- Hochberg, J. E. (1968). In the mind's eye. In R.N. Haber (Ed.), *Contemporary theory and research in visual perception* (pp. 309–331). New York: Holt, Rinehart & Winston.
- Itti, L. (2005). Models of bottom-up attention and saliency. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 576–582). San Diego, CA: Elsevier.
- , Rees, G., & Tsotsos, J. K. (Eds.). (2005). *Neurobiology of attention*. San Diego, CA: Elsevier.
- Julesz, B. (1984). A brief outline of the texton theory of human vision. *Trends in Neuroscience*, *7*, 41–45.
- Levin, D. T. (2002). Change blindness blindness as visual metacognition. *Journal of Consciousness Studies*, *9*, 111–130.
- Liddell, B. J., Brown, K. J., Kemp, A. H., Barton, B. J., Das, P., & Peduto, A., et al. (2005). A direct brainstem–amygdala–cortical “alarm” system for subliminal signals of fear. *NeuroImage*, *24*, 235–243.
- Mack, A., & Rock, I. (1998). *Inattention blindness*. Cambridge, MA: MIT Press.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
- McMains, S., & Somers, D. C. (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron*, *42*, 677–686.

- Merkle, P. M., & Daneman, M. (1998). Psychological investigations of unconscious perception. *Journal of Consciousness Studies*, 5, 5–18.
- , & Joordens, S. (1997). Parallels between perception without attention and perception without awareness. *Consciousness and Cognition*, 6, 219–236.
- Merkle, P., & Reingold, E. (1992). Measuring unconscious perceptual processes. In R. Bornstein & T. Pittman (Eds.), *Perception without awareness: Cognitive, clinical, and social perspectives* (pp. 55–80). New York: Guilford.
- Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.
- Neumann, O. (1990). Visual attention and action. In O. Neumann & W. Prinz (Eds.), *Relationships between perception and action: Current approaches* (pp. 227–267). Berlin: Springer.
- Norretranders, T. (1999). *The user illusion: Cutting consciousness down to size* (Chaps. 6, 7, 10). New York: Penguin Books.
- Oliva, A. (2005). Gist of a scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier.
- Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.
- Po, B. A., Fisher, B. D., & Booth, K. S. (2003). Pointing and visual feedback for spatial interaction in large-screen display environments. In *Proceedings of the 3rd International Symposium on Smart Graphics* (pp. 22–38).
- Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109, 160–174.
- Pylyshyn, Z. (2003). *Seeing and visualizing: It's not what you think*. Cambridge, MA: MIT Press.
- Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, 7, 17–42.
- . (2001). Change blindness: Implications for the nature of attention. In M. R. Jenkin & L. R. Harris (Eds.), *Vision and attention* (pp. 169–188). New York: Springer.
- . (2002a). Change detection. *Annual Review of Psychology*, 53, 245–277.
- . (2002b). Internal vs. external information in visual perception. *Proceedings of the Second International Symposium on Smart Graphics* (pp. 63–70).
- . (2003). Visual attention. In L. Nadel (Ed.), *Encyclopedia of cognitive science*. London: Nature Publishing Group.
- . (2004a). The invariance of visual search to geometric transformation. *Journal of Vision*, 4, 178a. <http://journalofvision.org/4/8/178>.
- . (2004b). Visual sensing without seeing. *Psychological Science*, 15, 27–32.
- , & Cavanagh, P. (2004). The influence of cast shadows on visual search. *Perception*, 33, 1339–1358.
- , & Enns, J. T. (1995). Preemption effects in visual search: Evidence for low-level grouping. *Psychological Review*, 102, 101–130.
- , & Enns, J. T. (1998). Early completion of occluded objects. *Vision Research*, 38, 2489–2505.
- , O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8, 368–373.
- Sharpe, S. H. (1998). *Conjurer's psychological secrets*. Calgary, Alberta, Canada: Hades.
- Simons, D. J., Nevarez, G., & Boot, W. R. (2005). Visual sensing is seeing: Why “mindsight,” in hindsight, is blind. *Psychological Science*, 16, 520–524.
- , & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, 9, 16–20.
- Tatler, B. W. (2002). What information survives saccades in the real world? In J. Hyönä, D. P. Munoz, W. Heide, & R. Radach (Eds.), *Progress in Brain Research*, 140, 149–163.
- Theeuwes, J., & Godijn, R. (2002). Irrelevant singletons capture attention: Evidence from inhibition of return. *Perception & Psychophysics*, 64, 764–770.
- Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology*, 40A, 201–237.
- , & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95, 15–48.
- Tsotsos, J. K. (1990). Analyzing vision at the complexity level. *Behavioral and Brain Sciences*, 13, 423–445.
- Ware, C. (2004). *Information visualization: Perception for design* (2nd ed.). San Francisco: Morgan Kaufman.
- Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience*, 18, 411–418.
- Wolfe, J. M. (1992). “Effortless” texture segmentation and “parallel” visual search are not the same thing. *Vision Research*, 32, 757–63.
- , Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 419–433.
- . 1999. Inattentional amnesia. In V. Coltheart (Ed.), *Fleeting memories* (pp. 71–94). Cambridge, MA: MIT Press.