

The Management of Visual Attention in Graphic Displays

Ronald A. Rensink

Departments of Computer Science and Psychology

University of British Columbia

Vancouver, Canada

Surface mail:

Ronald A. Rensink
Department of Computer Science
2366 Main Mall
Vancouver, BC
V6T 1Z4
Canada

Email: rensink@cs.ubc.ca; rensink@psych.ubc.ca

Website: <http://www.cs.ubc.ca/~rensink/>; <http://www.psych.ubc.ca/~rensink/>

Word count (abstract): 90

Word count (main body): 8882

In C. Roda (ed.), Human Attention in Digital Environments, Cambridge: University Press.

Abstract

This chapter presents an overview of several recent developments in vision science, and outlines some of their implications for the management of visual attention in graphic displays. These include ways of sending attention to the right item at the right time, techniques to improve attentional efficiency, and possibilities for offloading some of the processing typically done by attention onto nonattentional mechanisms. In addition it is argued that such techniques not only allow more effective use to be made of visual attention, but also open up new possibilities for human-machine interaction.

The Management of Visual Attention in Graphic Displays

Graphic displays such as maps, diagrams, and visual interfaces have long been used to present information in a form intended to be easy to comprehend (e.g., Massironi, 2002; Tufte, 2001; Ware, 2008). While it is clear that such a goal is important, it is not so clear that it has always been achieved. Are current displays for the most part *effective*—do they enable user performance to be rapid, easy, and accurate? Or are better designs possible?

These concerns are discussed here in the context of how to manage visual attention in graphic displays (including *visual* displays¹). This chapter is not directly concerned with the design of displays that respond effectively to the user (e.g., Roda and Thomas, 2006; Vertegaal, 2003). Rather, it focuses on the complementary perspective: how to design a display so that the user responds effectively to *it*. Results here apply equally well to static, dynamic, and interactive displays. For interactive displays the separation of the two perspectives need not be absolute: management of attention could depend on what the user does (e.g., the particular items highlighted could depend on where the user is looking). Indeed, designs that address both sets of issues—how the machine might effectively respond to the user *and* the user to the machine—will ultimately form the complete basis for effective interaction. But before tackling this, the best course is to first understand the individual perspectives separately. As such, discussion will be limited to the ways that graphic displays can support the management of attention in a user.

Another restriction is that this chapter focusses exclusively on *visual* attention—i.e., those attentional processes that allow us to see. It will not be greatly concerned with many issues associated with “attention” in general, e.g., task switching, or keeping attention on a particular task. The topic of attention is a highly complex one, with a great deal of associated literature; it cannot be covered in depth in a single chapter (for further information, see, e.g., Itti et al, 2005;

Parasuraman, 2000; Pashler, 1998). However, as “attention” is a term covering a set of processes that are largely independent of each other (Allport, 1993), it is possible to focus on those processes that pertain to the visual “picture” we experience. This includes not only eye movements, but also other, less visible processes. These processes often align in their operation, allowing “attention” to be treated as a single process. But occasionally the characteristics of individual components become relevant, requiring descriptions that are more specific.

With these caveats in place, the stage is set. Discussion begins with a brief overview of recent findings concerning the role played by attention in visual perception. The subsequent sections outline three general approaches to managing it. The first involves ways to send attention where it is needed when it is needed. The second focuses on techniques for improving the efficiency of attention itself, so that it “locks on” to information with minimal time and effort. The third involves the possibility of offloading of attention to other, nonattentional mechanisms that carry out operations of considerable sophistication, and that may form the basis of new possibilities for human-machine interaction.

0. VISUAL PERCEPTION

To design displays that effectively manage visual attention, it is important to have some idea of its nature: what it does, what it does not do, and how it is controlled in “normal” viewing. Four groups of processes are relevant here: (i) those that act prior to attention, providing the “raw materials” on which it operates, (ii) the mechanisms constituting attention itself, (iii) the mechanisms that operate concurrently with—and independently of—attention, and (iv) the processes that co-ordinate these to create the picture we have of our surroundings. These processes are only part of what is involved in visual perception; for further details, see, e.g., Palmer (1999). For discussion of how perceptual considerations in general can influence the

design of graphic displays see, e.g., Few (2004), MacEachren (1995), and Ware (2008).

0.1. Early Vision

When discussing how we see, a natural place to begin is with *rapid* visual processes—those that take place within the first few hundred milliseconds of viewing. These require little attention and are not greatly influenced by observer expectations; they are typically carried out automatically and without conscious awareness. An alternate starting point is the set of processes that act directly on inputs from the eye. These *low-level* processes are also highly automatic, and are highly parallel, operating concurrently in all parts of the visual field. The set of processes that are both rapid and low-level constitutes what is known as *early vision* (see, e.g., Marr, 1982; Rensink and Enns, 1998).

Early vision creates an array of elements upon which all subsequent processing—both attentional and nonattentional—is based. These *features* include color, motion, contrast, curvature, and orientation in the plane. Importantly, features are not just the basic “building blocks” of visual perception, but are also involved in the control of attention: a unique feature is generally *salient*, automatically drawing attention to itself, and thus “popping out” to a viewer almost immediately (see, e.g., Treisman, 1988; Wolfe, 2000). Since these features are largely determined before attention has had a chance to act, this level of processing is sometimes referred to as *preattentive* vision.

Although most features are simple, the structures they describe can be complex. For example, the length of a partially occluded figure is determined via the extent of the completed figure, and not just the visible parts (Rensink and Enns, 1998). This indicates a degree of *visual intelligence* at early levels—relatively sophisticated processing even in the absence of attention. Indeed, the output of the early visual system may be best characterized in terms of *proto-objects*

(localized precursors of objects) rather than simple measurements (Rensink and Enns, 1998; Rensink, 2000). Other structures also exist at this level. For example, arrays of items sharing similar features often form groups that extend over large regions of space (see, e.g., Pylyshyn, 2003, ch. 3; Wolfe, 2000).

Early processes can recover several scene-based properties, such as three-dimensional orientation, direction of lighting, surface convexity/concavity, and shadows. Such estimates are formed on the basis of “quick and dirty” assumptions about the environment that are true most of the time (Rensink and Cavanagh, 2004), in accord with the proposal that early vision produces a viewer-centred description of the world that is represented in a fragmented fashion (Marr, 1982).

0.2. Visual Attention

A key factor in much of vision (including the conscious picture we experience) is visual attention. Although at the subjective level we generally have no problem understanding what it means, attention has proven surprisingly resistant to an objective formulation. It has sometimes been viewed as a “filter” or “fuel” (see, e.g., Wickens and McCarley, 2008). But it can also be usefully characterized simply as selective control, carried out in different ways by different processes (Rensink, 2003). There appear to be several types of visual attention—including eye movements as well as several covert processes—which may or may not be directly related to each other. (For a set of current perspectives, see Itti et al., 2005). In all cases, these processes appear to be extremely limited in capacity, with only a few items attended at any time, and only a few properties of those items (e.g. Hayhoe et al., 1998). Some of the more commonly-encountered types are as follows:

0.2.1. Selective integration

One important type of attention is *selective integration*: the binding of selected parts or properties into a more complex structure. For example, searching for a single L-shaped item among several T-shaped items often takes a while; the distinguishing property—the arrangement of the horizontal and vertical segments—is evidently not salient. It likewise takes time to detect unique combinations of orientation and color, or more generally, combinations of most features. It has been proposed that the detection of such combinations is via a *spotlight of attention* that selectively integrates the features at each location into an *object file* at a rate of about 50 milliseconds per item (Treisman, 1988; Wolfe, 2000). Thus, if a target element has a salient feature, it will automatically pop out (Section 0.1); otherwise attention must travel around the display on an item-by-item basis until the target is integrated, and seen.

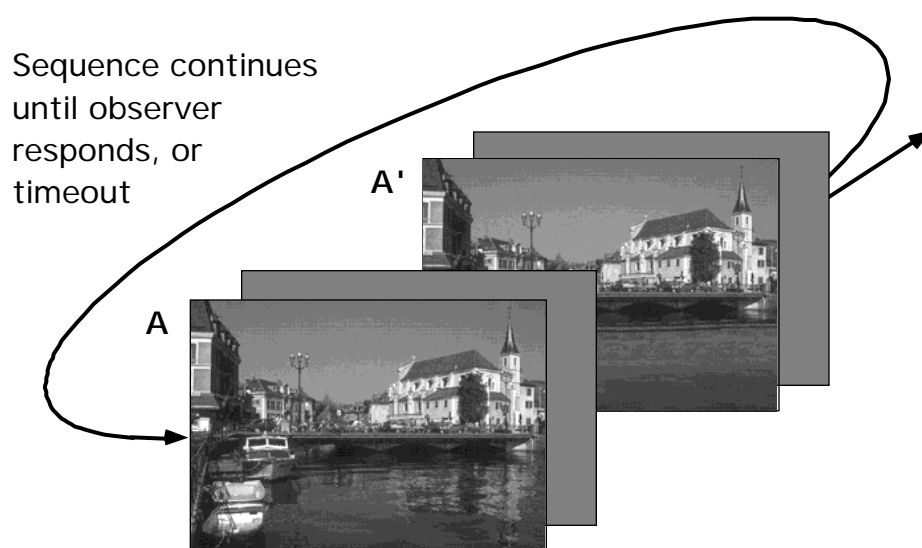


Figure 1. Flicker paradigm. Original image A (harbour with reflection of castle) and modified image A' (harbour with reflection gone) are displayed in the sequence A, A', A, A', ... with gray fields between successive images. Observers typically require several seconds to see such a change, even though it is large and easily seen once noticed.

0.2.2. *Selective coherence*

Figure 1 shows an example of the *flicker paradigm*, where an original image A continually alternates with a modified image A', with brief blank fields between successive images. Under these conditions, observers usually have great difficulty noticing changes, even when the changes are large, repeatedly made, and expected by the observer. Indeed, this *change blindness* can exist for long stretches of time—up to 50 seconds under some conditions (Rensink et al., 1997).

Change blindness can be accounted for by the proposal that attention is needed to consciously see change—i.e., it selectively integrates structures across time as well as space (cf. Section 0.2.1). A change will be difficult to notice if its accompanying motion transients do not draw attention to its location—e.g., if they are swamped by other motion signals in the image. More generally, change blindness occurs whenever such swamping takes place, such as if the change is made during an eye movement, eyeblink, or occlusion by some other object (Rensink, 2002a).

The perception of change can be explained by *coherence theory* (Rensink, 2000). Prior to attention, proto-objects are continually formed in parallel across the visual field (Figure 2). Attention selects a few of these for entry into a *coherence field*, a circuit between the attended items and a single, higher-level *nexus*; this can be viewed as the holding of information in visual short-term memory. The proto-objects are thus “knit” into a representation with spatio-temporal coherence, corresponding to a single coherent object. Attention is released by breaking this circuit, with the object dissolving back into its constituent proto-objects. There is little after-effect of having been attended, with no accumulation of items in conscious visual experience (also see Wolfe, 1999).

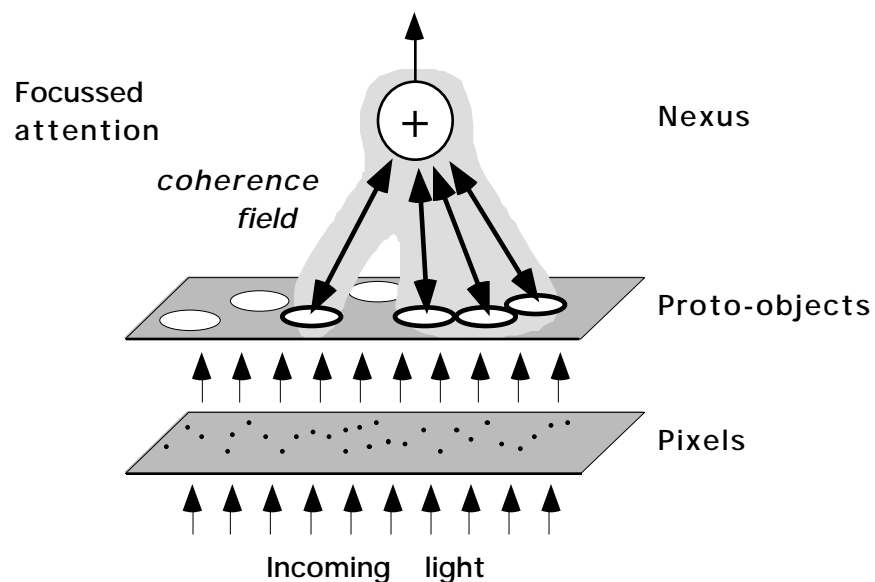


Figure 2. Coherence theory. Early vision continually creates proto-objects across the visual field. These are *volatile*, lasting only a short time; if a new item appears at the location of an existing item, its representation simply overwrites the old one. Attention can select a few proto-objects and set up reciprocal connections between them and an attentional nexus, resulting in a *coherence field*. (The “+” is purely symbolic, simply indicating that some form of pooling is used.) As long as the proto-objects are held in this field, they form a single object with both temporal and spatial coherence (Rensink, 2000).

In this view, attended items are not entirely independent, but rather are parts of a coordinated complex: information is pooled into a single nexus, perhaps by taking the sum or the maximum of all inputs. It is therefore difficult for the nexus to differentiate between receiving a single change signal or multiple change signals, since the nonlinear nature of the pooling leads to largely to the same result in both cases. Consequently, observers can never distinguish two changes at a time from one, no matter how hard they try (Rensink, 2001, 2002a).

0.2.3. Selective experience

Recent studies also indicate that attention is needed not just to see change, but even to simply *see* (i.e., to have a conscious picture of an element). For example, Mack and Rock (1998) asked observers to view an overlapping pair of lines (one horizontal and one vertical), and judge

which was longer. After several such trials, a display was presented containing an unexpected test item (Figure 3). Observers often failed to see the test item, even when they looked at it directly. Such *inattention blindness* can occur even for objects that are highly visible, such as a person in a gorilla suit walking across a scene (Simons and Chabris, 1999).

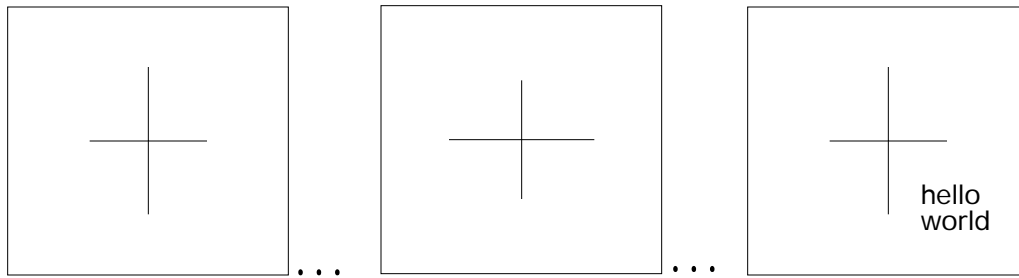


Figure 3. Inattention blindness. Observers are shown a sequence of images, and asked to determine which line (horizontal or vertical) is longest in each. After several presentations, an image is presented containing an unexpected test stimulus. Observers often do not see this, even when looking directly at it (Mack and Rock, 1998).

Interestingly, objects that are not consciously seen (and thus, are not attended) can still influence perception. For example, unseen lines surrounding a test item have been found to induce a length illusion in that item (Moore and Egeth, 1997). This reinforces the findings from early vision, which point towards a form of visual intelligence that involves neither conscious awareness nor attention.

0.3. Nonattentional Processing

A common intuition about vision is that it exists entirely to produce a sensory experience of some kind (i.e., a picture) and that attention is the “central gateway” for doing this. However, there is increasing evidence that several kinds of sophisticated processing can be done without attention, and that some of these processes have nothing directly to do with visual experience

(see, e.g., Rensink, 2007). These processes are believed to be based on early vision and operate concurrently with attentional processes, but independently of them.

0.3.1. *Statistical summaries*

An interesting form of visual intelligence is the ability to rapidly form *statistical summaries* of sets of briefly-presented items. For example, when observers are briefly presented with a group of disks, they can match the mean size of these to an individual disk as accurately as they can match two individual disks (Ariely, 2001). This can be done using exposures of as little as 50 milliseconds (Chong and Treisman, 2003), indicating that attention is not central to the creation of such summaries. This ability may also extend to other statistical measures, such as range or variance, although this has not yet been confirmed.

0.3.2. *Scene gist*

Evidence is also accumulating that several aspects of scene composition can be determined rapidly and without attention. One of these is its abstract meaning (or *gist*)—e.g., whether it is a city, kitchen, or farm. Gist can be determined within 150 milliseconds, a time insufficient for attending to more than a few items (Thorpe et al., 1996). It can be reliably extracted from blurred images, indicating that details are not important. Interestingly, the representation of two different gists can be activated simultaneously (Oliva, 2005).

It also appears possible to determine several related properties this way. For example, observers can also rapidly determine how open or crowded a scene is (see, e.g., Oliva, 2005). All these are likely based on the distribution of early-level features in the image (e.g., line orientations or colors), possibly by relying on rapid statistical summaries. They do not appear to involve coherent object representations.

0.3.3. *Scene layout*

Another possible nonattentional process is memory for *layout*—the spatial arrangement of objects in the scene (Hochberg, 1978). This may also contain a small amount of featural information, such as coarse descriptions of their size, color, and orientation. Some layout information may be extracted within several seconds of viewing—likely via eye movements or attentional shifts—and can be maintained over intervals of several seconds without the need for attention (Tatler, 2002). Interestingly, memory for repeated layouts can be formed in the absence of awareness that such patterns are being repeated (Chun and Jiang, 1998).

0.3.4. *Visuomotor guidance*

It has been proposed (Milner and Goodale, 1995) that vision involves two largely separate systems: a fast, nonattentional *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 conscious perception and the 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.

0.4. Scene Perception

The discovery that attention is needed for conscious visual experience has several counterintuitive implications. For example, since attention has a limited capacity (e.g., Pylyshyn and Storm, 1988), only a few objects can be consciously seen at any time. But if this were the case, why do we not notice such limitations? Why do we believe we see all objects and all events, and each of these in great detail?

The answer to these involves a shift in our view of how information is integrated to perceive a scene. Originally, it was believed that such integration involved the creation of a dense, static representation, in accord with our impressions as observers. But more recent work is beginning to view integration in terms of dynamic rather than static processes—in terms of *co-ordination* rather than *construction* (see Ballard et al., 1997; Rensink, 2007).

0.4.1. *Virtual representation*

The seeming contradiction between our impression of virtually unlimited perceptual content and the existence of severe attentional limitations can be accounted for by the idea of a *virtual representation*: instead of forming a coherent, detailed representation of all the objects in our surroundings, the visual system only creates a coherent representation of the items needed for the task at hand (Rensink, 2000). If attention can be managed such that a coherent representation of an object can be created whenever needed, the scene representation will appear to higher levels as if all objects and events are represented simultaneously. Such a representation would have all the power of a “real” one, while using much less in the way of processing and memory resources.

In this view, the conscious seeing of a display relies on a relatively sparse, dynamic just-in-time system that creates coherent representations whenever these are needed. Among other things, this implies that there is little or no general-purpose representation in vision: whatever is used at any moment is coupled to the task at hand, and would likely be suboptimal for other purposes. Different people will literally see a scene in different ways, depending on their individual knowledge and expectations (Rensink, 2007).

0.4.2. *Triadic architecture*

The successful use of virtual representation requires that attentional shifts be made to the appropriate items at the appropriate times. How might this be implemented in the human visual system? One possibility is a *triadic* architecture (Rensink, 2000), sketched in Figure 4. As its name implies, this architecture is based on three interacting subsystems:

1. *Early visual system.* This rapidly creates proto-objects in parallel across the visual field; these are volatile, and must be continually regenerated. This system operates automatically and continually, without the awareness of the viewer.
2. *Attentional system.* This can form a set of selected proto-objects (no more than 3-4) into a coherent object representation. This is the basis of the conscious perception of change, and possibly, conscious perception in general.
3. *Setting system.* This provides a coarse description of the scene to help guide attention to appropriate items. It is largely based on properties such as gist and layout, which can be obtained—or at least maintained—without attention. These can invoke knowledge in long-term memory, which in turn constrains expectations as to likely objects, likely actions, etc. Although some information can be accumulated in the setting system, it is used for guidance only, and does not form part of the picture that is experienced.

These largely correspond to the groupings described in previous sections—early vision, visual attention, and nonattentional processing—except that the setting system contains only those nonattentional processes that control attention. (Those that guide visuomotor actions, for example, belong to a separate group.) The connection to long-term knowledge helps inform attentional control, likely by acting through the layout system, and perhaps also by affecting

attentional content directly. Most of long-term memory is not in play at any instant, and so is not considered part of here-and-now visual perception.

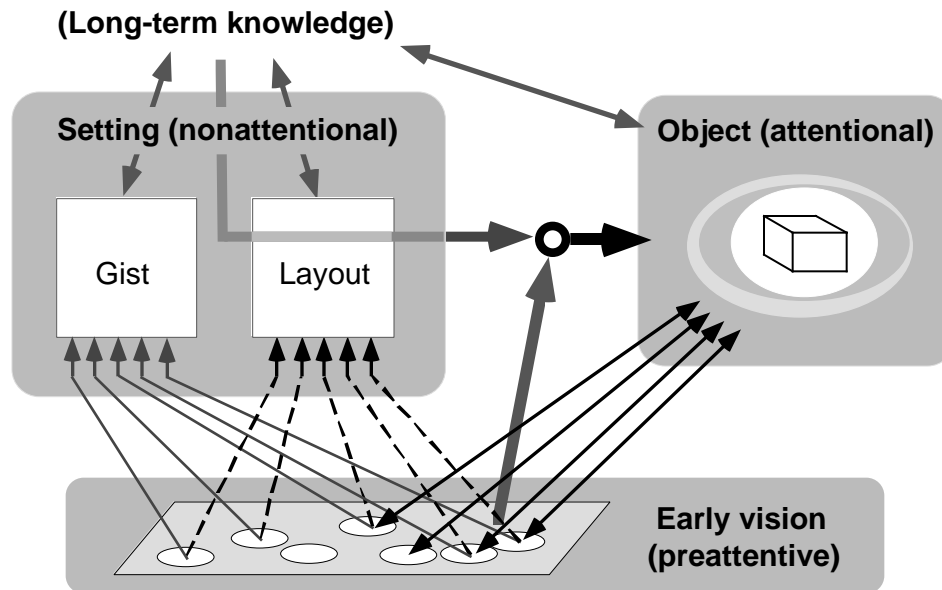


Figure 4. Triadic architecture (Rensink, 2000). Perception is carried out via three interacting systems. (1) Early-level processes create volatile proto-objects. (2) Visual attention "grabs" selected structures and forms them into an object with both temporal and spatial coherence. (3) Setting information (obtained via nonattentive processes) guides the allocation of visual attention on the basis of long-term knowledge. Control is based on the combination of this with the automatic drawing of attention based on physical properties at early levels.

In this view, then, the perception of a scene (or a display) occurs via the constant allocation of visual attention, which is largely controlled by two mechanisms: (i) bottom-up drawing of attention to salient items on the basis of their low-level physical properties, and (ii) top-down guidance to important items on the basis of high-level knowledge about the objects and the scene. Thus, what is seen reflects a sophisticated balance between the immediate, ever-changing information from the world, and the longer-term, more stable knowledge of the viewer.

1. COERCION

The view of perception emerging from recent work in vision science is based on a just-in-time co-ordination of several subsystems, one of which is visual attention (Section 0.4). A critical part of this co-ordination is the effective allocation of attention, so that it arrives at the right place at the right time.

Given that the visual experience of an observer depends on the co-ordination of attention, and given that this co-ordination is strongly affected by what is shown to the eyes, the possibility arises of *coercive graphics*: displays that coerce² attention to make the observer see (or not see) a particular part of a display in a natural way (Rensink, 2002b, 2007). In essence, the mechanisms that manage attention in everyday seeing are “hijacked” to control the viewing experience of the user. Such coercion has long been used by magicians and filmmakers to achieve a variety of striking effects, most of which are incorporated seamlessly into the experience of the viewer (e.g., Kuhn et al., 2008). Given machines with “superhuman” control over what is displayed, the potential exists for coercive displays that are even more powerful than these.

In any event, successful coercion could result in an observer always attending to whatever was needed by the task at hand. Apart from a general improvement of performance, such coercion might also be useful for specialized populations. For example, it might assist those users who have difficulty sending their attention to the right item at the right time; among these might be new users of a system, who often do not know what to attend, or when. Another potential application would be *soft alerts* that would not disturb existing attentional control (Rensink, 2002b, 2007). Such alerts would be particularly useful for situations where the arrival

of a new event does not require immediate attention—e.g., the arrival of email while the operator is monitoring some unrelated task (e.g., McCrickard and Chewar, 2003).

Several different types of attentional coercion appear to be possible. Each of these involves a different set of mechanisms, and has its own strengths and weaknesses.

1.1. Low-level Properties

In everyday seeing, attention is automatically—and usually involuntarily—*drawn* to particular items or locations based on their physical properties (Section 0.1). Much of this is based on *saliency*, a quantity that governs the priority of attentional allocation: the higher the saliency, the more likely attention is to be drawn. Such control is thought to be largely independent of the beliefs and goals of the viewer, although some aspects may be affected by the task and instruction set (Egeth and Yantis, 1997; Theeuwes and Godijn, 2002). Thus, a considerable amount of coercion can be achieved simply by *highlighting* the target item (or region) to enhance its saliency. At least five possible ways exist of doing this.

1.1.1. Featural cues

Saliency is largely based on differences in the density of features in a region, with large differences creating the highest levels of saliency. Thus, attention can be coerced by the use of *featural cues* in a display: if a unique feature exists in an area, saliency will be high, and the corresponding item will simply “pop out”. Even if its saliency is somewhat lower (e.g., if the difference in features from its surround is not that great), an item will still be attended relatively quickly (Itti, 2005). Properties that can be used for this include:

- brightness (contrast)
- colour (hue)
- length
- width

- orientation (2D)
- curvature (2D)
- convexity/concavity (3D)
- motion
- blinking

Importantly, only simple properties appear to exert such control. For most of these (e.g., brightness, length), the absolute level of the features is important, with high levels more effective than low ones; for others (e.g., orientation), the degree of difference is also a factor (Figure 5). For details, see, e.g., Ware (2004, ch. 5), Wickens and McCarley (2008), and Wolfe (2005).

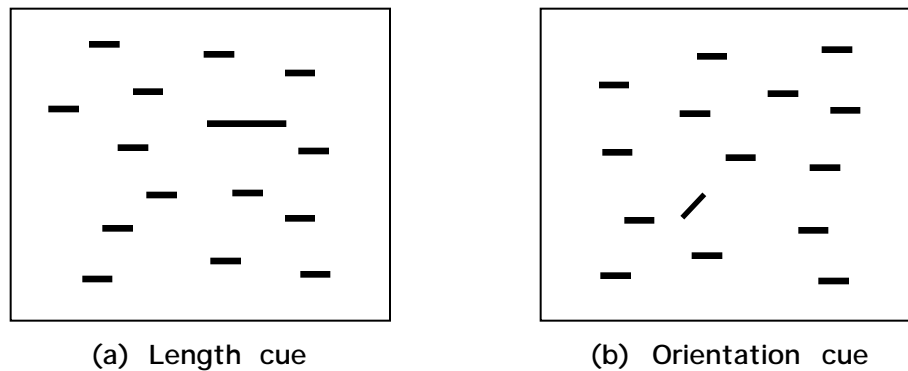


Figure 5. Featural cues. Attention is drawn to items with features differing from those of their neighbours. (a) Unique length. Attention is drawn to the item that is longer than its neighbours; a shorter item among longer ones would not be as salient (see, e.g., Wolfe, 2000). (b) Unique orientation. Attention is drawn to the item for which orientation is different. While there may be a difference in the categories, the *degree* of the orientation difference is also important.

It is important to note that saliency is based upon the properties of relatively complex proto-objects, and not simply pixels in the image (Section 0.1). For example, a graphical element with a distinctive size or orientation will not be salient if it is part a proto-object that is itself undistinguished (Figure 6).

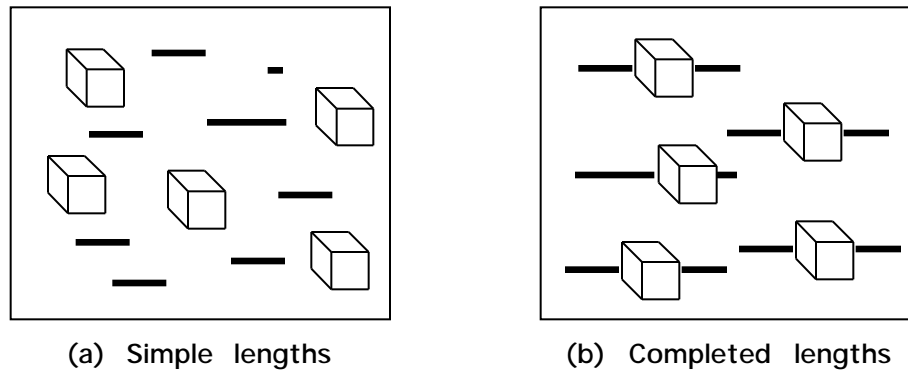


Figure 6. Proto-object structure (Rensink and Enns, 1998). (a) A unique simple length pops out if distinct from others in the region. (b) When it is incorporated into a proto-object (here, the completed bar), it becomes unavailable, even though the same pixels are in the image as before. The proto-objects (bars) are not distinct, and so attention is not drawn to them automatically.

1.1.2. Lighting level

Low-level coercion can also be achieved by a literal highlighting of an item in the depicted scene, with a more brightly-lit region drawing attention (Khan et al., 2005; Seif El-Nasr, 2005). It is likely that the absolute luminance of the region is not important here; rather the key factor may be the perceived level of illumination. Although this has not yet been verified in controlled experiments, it is known that lighting can be rapidly separated from surface brightness at early levels (Section 0.1).

Conversely, it might also be possible to use lighting to divert attention away from a particular region or object via lighting. For example, given that shadows are a natural opposite to highlights, and given that they can also be identified at early levels (Rensink and Cavanagh, 2004), it might be possible to keep attention away from items or regions perceived as being in the shadows.

1.1.3. *Level of detail*

Another approach to highlighting involves the *level of detail* in an item or region. Here, coercion is based on the fact that saliency is high for items with high levels of detail, which contain relatively high spatial frequencies (Itti, 2005). Attention can therefore be coerced to a given item by using (selective) blurring to remove high-frequency components or features in the rest of the image, which then increases the saliency of the target item without affecting its appearance (Kosara et al., 2002; Su et al., 2005). Possibly related to this is the finding that viewers prefer items and regions with a greater level of detail, even when the differences in detail are not consciously noticed (Halper et al, 2003; see also DeCarlo and Santella, 2002).

1.1.4. *Configural focus*

Attention can also be drawn to a location via configurations that extend over a region of space (Figure 7). For example, attention is automatically drawn to the intersection point of two or more lines that are sufficiently long (Ward, 1996); this does not occur if the lines are short (Wolfe and DiMase, 2003). Attention is also drawn to the vanishing point of lines in one-point perspective (Ward, 1996). Both effects may be based on the fact that such a vanishing point corresponds to the viewer's direction of gaze.

Interestingly, the movement of random dots away from a single point—its *focus of expansion*—over a large area attracts attention; no other motion patterns induce such behaviour (von Mühlenen and Lleras, 2007). This is likely because the expansion corresponds to the looming of an object travelling directly towards the viewer, which requires immediate response (Figure 7). Expansion of an item over a small area does not appear to exhibit such an effect (Bartram et al, 2001), again indicating the importance of extended structure.

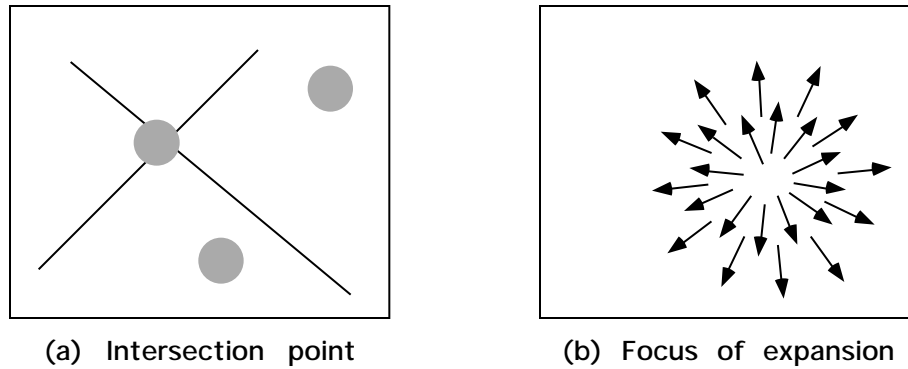


Figure 7. Drawing of attention via configural focus. (a) Intersection point of a pair of extended lines. Attention is automatically drawn to this point (or any items at that location), possibly because of its similarity to a vanishing point. (b) Focus of expansion over an extended area. Attention is automatically drawn to the point from which random dots move, but only if their movement is outward, corresponding to an object travelling towards the viewer. In some ways, this could be considered the dynamic analog to the intersection point.

Such configurations could highlight an item if placed in the background (ideally, at low contrast), with the item being at the “focus”. The computation of such structure—both static and dynamic—likely involves processes that act rapidly over large areas (Section 0.3). If so, such highlighting could take place rapidly, and without interfering with other aspects of perception.

1.1.5. Centres of gravity

An item attracts more attention when placed at the centre of a display—or more precisely, when at the centre of gravity of the elements that the display contains (Bang, 2000; Solso, 1994). Such increased saliency might also exist for centres of gravity of individual objects or groups, given that these appear to be the bases of eye movements and other perceptual measures (Vishwanath and Kowler, 2003). Consistent with this, when tracking an item, attention appears to be concentrated around its centre of gravity (Alvarez and Scholl, 2005). Taken together, these results suggest the existence of a general strategy of attentional control based on centres of gravity, possibly applicable to all levels of organization, and carried out via rapid nonattentional processes (Section 0.3).

1.2. High-level Interest

A rather different set of techniques for attentional coercion involves the voluntary *direction* of attention to an item. This is governed by more abstract, higher-level factors such as the viewer's interest in a particular object; these factors are generally contingent, depending on the particular viewer and the particular task carried out. This form of control is slower and more effortful than low-level control, and involves different mechanisms. The exact way these two types of control interact has not yet been established. However, high-level control can override low-level considerations if the interest in a particular item is sufficiently high (see Egeth and Yantis, 1997).

Although “interest” has a reasonably clear subjective meaning (e.g., Rensink et al., 1997), it is difficult to give it an objective formulation. Some aspects can be captured in terms of features that are unexpected in the context of the given scene (Bruce and Tsotsos, 2006; Elazary and Itti, 2008). But such a characterization is necessarily incomplete—if nothing else, there is always a possible dependence on the task. More generally, interest must involve a high-level awareness of the situation at hand, and the mechanisms of such awareness are only beginning to be understood (e.g., Endsley, 1995). In the absence of a firm theoretical framework, it may be best to design displays based on practical considerations obtained from other domains. For example, interest in an item could be created via the techniques of showmanship used by magicians (Kuhn et al., 2008; Sharpe, 1988; Tognazzini, 1993). The idea of computer interface as theatre (Laurel, 1993) may also be relevant, with interactions viewed as parts of a larger-scale situation which emphasizes some aspects of an interface, and de-emphasizes others.

In any event, designs can be quantifiably tested via techniques such as the flicker paradigm (Section 0.2). For example, given that attention is needed to see a change, objects that

are seen to change relatively quickly under these conditions can be interpreted as more quickly attended, and thus, more interesting. Systematic testing might provide an ordering of the items in the display in terms of their interest for a given observer, or a given task. Indeed, such techniques could be used to determine which particular parts or structural levels are most interesting in an image (New et al., 2007; Rensink, 2002b).

1.3. Learned Associations

Another potentially useful set of controls are *learned associations*. These can be viewed as hybrids of low- and high-level mechanisms: simple words and shapes that control attention quickly and automatically (like low-level control) on the basis of their meaning (like high-level control). Since the formation of these associations depends on learning, there is a possibility (yet untested) that they may not be universally effective. Instead, they may depend on the cultural background—and perhaps even the individual history—of the viewer.

In any event, this kind of control can be quite powerful if used correctly (Kuhn et al., 2008). Two different types can be distinguished, based on their function:

1.3.1. Attractors

These are shapes or words that draw attention to themselves on the basis of their significance or meaning (i.e, semantics). For example, it is believed that attention is drawn automatically to the eyes of a human face (Block, 2001). It is also drawn automatically to a person's name, or to any other emotionally-laden word or symbol; the meaning of these can apparently be obtained without attention (Mack and Rock, 1998, ch. 6).

1.3.2. Directives

These are shapes or words that automatically send attention to a location other than their own (Rensink, 2002b). For example, if a viewer attends to an image (or depiction) of a person in a display, their attention will automatically follow that person's gaze. Likewise, an image of a pointing finger will direct a viewer's attention to the item being pointed at (see, e.g., Burton et al., 2009; Kuhn et al., 2008).

All of these likely reflect learned responses to social cues, which are typically of great importance to humans. However, this kind of direction can also be induced by schematic figures such as arrows, although the effects of these are somewhat weaker (Ristic et al., 2007).

1.4. Unseen Coercion

A potentially important—although currently speculative—type of control is based on graphical elements that do not enter into the conscious experience of the user, but still keep their coercive power. This type of coercion could form the basis of “magical” displays in which a viewer would be guaranteed to see (or not see) a selected item, while experiencing nothing out of the ordinary (Rensink, 2002b).

Such coercion could be done—at least in theory—by elements that are unattended. These can remain unseen while still having significant effects on various aspects of perception, including attentional control (e.g., Mack and Rock, 1998). The feasibility of this approach therefore depends on the extent to which this can be done. One possibility is to present elements so briefly they are not consciously registered (e.g., Marcel, 1983); another is to draw attention away by a distractor, with the coercing items appearing in some other location where attention is necessarily absent. Indeed, once the first unseen coercion is accomplished a “coercive cascade”

might begin, with each coercive element providing enough attentional diversion to allow the subsequent coercive element to remain unseen.

2. EFFICIENCY

An important part of the management of attention is to ensure that it gets to the right place at the right time. But it is also important to ensure that minimal effort is expended in doing so, that few errors are made, and that good use is made of attention when it arrives at its destination. In other words, it is important that *attentional efficiency* be high³. Otherwise the user might be needlessly slowed down or fatigued. Indeed, in the extreme case, the user might be asked to do something that visual attention simply cannot carry out.

Given that attention is managed by a variety of systems (Section 0.4), it follows that there exist various ways of improving its efficiency. Three of the more important ones are as follows.

2.1. Low Saliency

Ideally, attention should arrive at the desired location as quickly as possible, with minimal expenditure of effort. In normal viewing, it is drawn automatically and involuntarily to locations of high saliency (Sections 0.1, 1.1). But while saliency is often useful in attracting attention to an item or location, it also has a dark side in that salient locations will draw attention regardless of the situation at hand. If the saliency of an irrelevant item is sufficiently high, it will therefore cause distraction, drawing attention away from where it should go, and requiring the viewer to devote time and effort to overcoming these effects.

A useful way of achieving high efficiency of allocation is therefore to keep the saliency of nonessential graphical elements as low as possible. One way of doing this is to require that nonessential elements not vary greatly—or sharply—in the values of the features they use (for a

partial list of such features, see Section 1.1). While these features need not be absolutely uniform, unique values and sudden changes should be minimized. In particular, the use of motion in a display should be avoided whenever possible, since it is a particularly effective feature for attracting attention (e.g. Bartram et al, 2001; Ware, 2004, ch. 5),

Because elements that are nonessential in one task may be essential in another, it is important to keep in mind that low saliency could be achieved for different elements at different times. Thus, for example, displays could have the properties of their graphical elements depend on the particular mode of use, with only the essential elements being salient at any instant.

2.2. Perceptual Organization

Another way of improving efficiency is via the use of perceptual organization. This can help in several ways, including limiting the number of items that attention can potentially act upon at any time, as well as organizing these into effective units of selection.

2.2.1. Reduced Clutter

The efficiency of attention can be improved simply by reducing clutter: given that attention is moved around on an item-by-item basis (Section 0.2), the fewer the items in a display, the faster and easier its allocation. Clutter has historically been associated with simple image-based properties, being expressed in measures such as the ratio of data to ink (Tufte, 2001, ch. 4-6). In this view, the main route to reducing clutter is to minimize the number of irrelevant items in the display, and the saliency of the remainder (e.g., St. John et al., 2005; Wickens and McCarley, 2008, pp. 76-77).

However, recent work in vision science shows that attention acts not on simple image properties but on more sophisticated early-level structures, such as proto-objects and extended groups (Section 0.1). Perceptual organization carried out at early levels can therefore help

reduce clutter further. For example, if the graphical elements in a display are positioned to form early-level groups, the number of effective items—and thus, the amount of clutter—is greatly reduced, even though the amount of “ink” is the same (Figure 8)

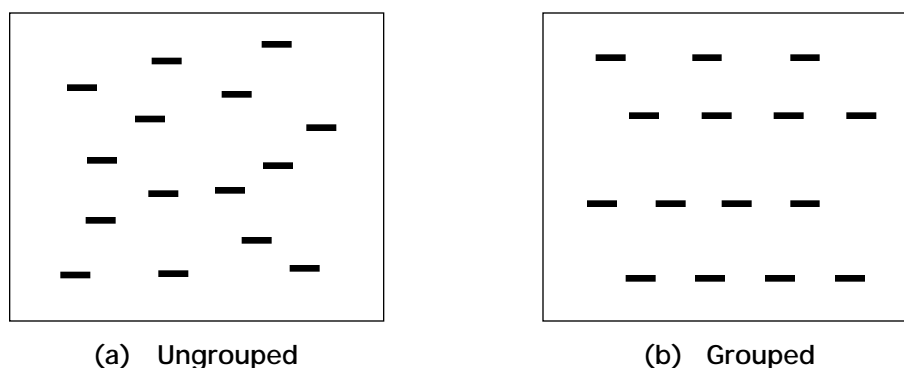


Figure 8. Reduction of clutter via grouping. (a) Ungrouped elements. Each element is a separate possibility for attentional allocation. The number of possibilities is high, and so clutter is considerable. (b) Grouped elements. Elements are immediately grouped on the basis of common alignment. Since attention is initially allocated to such groups, the number of possible operands is low, thereby reducing clutter.

2.2.2. Improved Selectivity

In addition to reducing clutter, grouping can also organize data to improve the ease and effectiveness of attentional selection. Early-level groups can be formed on the basis of several considerations, such as texture boundaries, alignment of elements, and element similarity (see, e.g., Wolfe, 2000). Once created, these become the effective operands of attentional allocation, with easy suppression of items in other groups (see MacEachren, 1995; Ware, 2004, ch. 6). Related to these are *surfaces* (Nakayama et al., 1995), two-dimensional “sheets” which facilitate the travel of attention between items (Figure 9).

Indeed, perceptual organization can go even further, forming *layers*: overlapping groups or surfaces that remain separate in regards to attentional operations (Mullet and Sano, 1995; Tufte, 1990). This can be done via transparent surfaces, with the elements of each layer on a separate surface (Mullet and Sano, 1995; Ware, 2004, ch.6); elements with common contrast

(colour) and scale, for example, can form such layers (Figure 9). Because they are overlapping, layers are useful in facilitating the allocation of attention in small-screen displays (Kamba et al., 1996). They have also been found useful for minimizing interference from reference structures such as grids, labels, and contour lines (Bartram and Stone, 2007).

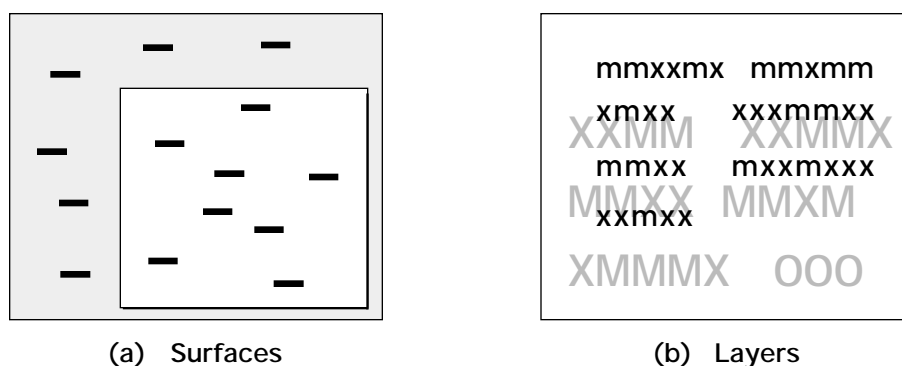


Figure 9. Organizational structures. (a) Surfaces. Attention travels easily between items on the same surface (i.e., bounded two-dimensional structure); it takes more time to travel between items on different ones. (b) Layers. Items are grouped on the basis of contrast and size to form two overlapping layers. When attention shifts among the items in a selected layer, there is relatively little interference from other items, even though they are in nearby locations.

2.3. Compatibility with Attentional Mechanisms

Recent research on human vision has found that relatively little of a scene is attended at any moment. Only a few items—and only a few properties of those items—can be attended at a given time (Section 0.2). If attention is to be used efficiently, it is therefore important for display design to be compatible with the attentional mechanisms used. In particular, information should be presented such that the limited capacity of these mechanisms is not strained.

Attentional mechanisms can always extract information whenever the corresponding graphical properties are perceptually distinct. But if efficiency is to be high, the amount of processing must be minimized. If the process of extracting information from a display can avoid the need for several passes at an element, and the need for complex processing to encode the

information contained in it, attention can “lock on” to the displayed information with minimal time and effort. This can be achieved—at least in part—via careful restrictions on the design.

2.3.1. *Restricted set of values*

One way to achieve high efficiency is for the properties of the graphical elements to match the “basic codes” of visual attention, so that distinct properties are represented by distinct mechanisms. This can be achieved by using a restricted set of values, which can—if carefully chosen—eliminate the need for sophisticated processing to make (and encode) any finer-grained distinctions that are unnecessary. Attentional pickup of information can then be as effortless as possible (see, e.g., Few, 2004; Ware, 2004).

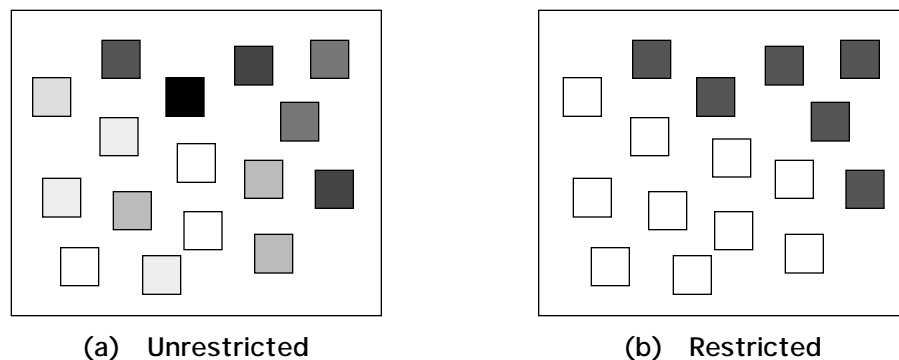


Figure 10. Effect of different sets of values. (a) Unrestricted set. Here, elements can take on a wide range of possible values. Distinguishing these from each other requires time and effort (attention). (b) Restricted set. Items now take on one of just a few possible values; the resultant display takes less time and effort to understand.

Although our understanding of this issue is far from complete, it appears that many of the features that attract attention are also basic encoding units (see Ware, 2008; Wolfe, 2005). In terms of commonly-used properties, about 4 distinct values (or 2 bits) appear to exist for each spatial dimension (size, orientation, etc), and about 8 basic values (3 bits) for color (see Healey, 1996; Shive and Francis, 2008; Ware, 2004, pp. 182-183; Wolfe, 2005). These could be used, for example, to represent particular ranges of size or particular categories. Finer-grained

distinctions should be avoided, since these would require a greater number of bits, and thus, more complex computations (Figure 10).

2.3.2. *Restricted positioning*

Another important approach to efficiency involves location—in particular, the distance between important graphical elements. If these are too close together (within $0.5E$ degrees of visual angle, where E is the distance from the centre of viewing), their locations will not correspond to separate position codes. The result is “crowding”, where the properties of individual items are no longer easy to access (Intriligator and Cavanagh, 2001; Wolfe, 2005). For high efficiency, therefore, adjacent elements must be sufficiently separated. A restricted set of locations would be one way of achieving this.

A second source of restriction stems from the need to minimize change blindness (Section 0.2). Change blindness can be induced by making the change during an eye movement (Grimes, 1996). Since eye movements make up about 10% of total viewing time on average, any transition will have at least a 10% chance of being missed. One way to lower this probability is by minimizing the need for (or the size of) eye movements—e.g., by restricting the separation of important information sources so that they are always close together.

2.3.3. *Restricted number of elements*

Another important aspect of compatibility is to ensure that the number of items used at any time does not exceed attentional capacity. Otherwise, a considerable amount of processing may be required to compensate, if it can be compensated at all.

For example, only about 4 moving items can be tracked at the same time (Pylyshyn and Storm, 1988; Pylyshyn, 2003, ch. 5); attempting to track more will inevitably result in some being lost. Indeed, for most tasks attentional capacity is about 4 items. The exact value differs

for different observers, but is usually around 3-5 items. A limit of 3 is a reasonable restriction for displays intended for most users; if more items need to be attended at any time, performance will tend to deteriorate.

An even more severe restriction applies to information that is conveyed dynamically (e.g., Albrecht-Buehler et al, 2005; Blok, 2006). Since attended information is pooled in a single nexus (Section 0.2), a user attending to two or more simultaneous changes will not be able to tell how many occurred, or which input was the source of a single change, even if given enough time (Rensink, 2001, 2002b). In such a situation, therefore, displays should have only a single information source in operation at any time.

3. OFFLOADING

Among the more interesting possibilities opened up by new findings on human vision is the offloading of attentional processing onto nonattentional systems. A common intuition about visual perception is that it is “attento-centric”—i.e., that attention is required for all important aspects of its operation. However, it is becoming increasingly apparent that perception is based on several systems, many of which operate concurrently with attention, and even independently of it (Section 0). Some of these have a high degree of visual intelligence, containing processes that—although limited in scope—show considerable sophistication (Sections 0.1, 0.3). Thus, if a display can be appropriately designed, a task traditionally done by “high-level”, attention-intensive processes could be offloaded onto faster, less effortful, and perhaps even more capable nonattentional systems, freeing up attention for other things (Card et al., 1999; Rensink, 2002b).

Because current knowledge of nonattentional perception is still far from complete, the full extent to which offloading can be usefully applied is unclear. However, it appears applicable to at least three general kinds of task.

3.1. Pattern Detection

Perhaps the best-developed example of offloading is the use of rapid nonattentional processes at early levels to detect patterns that would otherwise require conscious application of attention—i.e., “using vision to think” (Card et al., 1999). Here, numerical data is represented by simple graphical elements, with the properties of these elements coding the quantities of interest. Much of the focus of work in information visualization is on finding representations that allow the most suitable nonattentional mechanisms to be brought into play (e.g., Tufte, 2001; Ware, 2004). Important tasks here include detection of trends and detection of outliers. Both can be facilitated by rapid grouping at early visual levels (Section 0.1): the shape and density of the groups provides a visual representation of trends, while the grouping itself causes non-grouped items (outliers) to become salient, and thus, easily seen.

This approach can be extended beyond the analysis of pure numerical quantities. For example, complex multidimensional data tied to particular spatial co-ordinates can be usefully analyzed via carefully-designed maps (e.g., MacEachren, 1995). And the emerging area of visual analytics is heavily based on the offloading of high-level, attention-demanding analysis processes onto faster, nonattentional visual mechanisms (e.g., Thomas and Cook, 2005).

3.2. Statistical Estimation

One of the more interesting forms of visual intelligence is the ability to rapidly estimate the mean size of a briefly-presented set of items (Section 0.3). The speed of this process (within 50 milliseconds) suggests that it is carried out in the absence of attention. This ability may also

extend to other statistical measures—such as range or variance—although this has not yet been confirmed. Properties other than size might be also estimated this way; if so, the set of possible properties would likely include the features at early visual levels (Section 0.1).

Possibly related to this is the finding that viewers can rapidly estimate percentages in large sets of items, based on the preattentive visual features of colour (hue) and orientation (Healey et al., 1996). Estimates of absolute numbers can also be done rapidly, with speed depending only on the level of precision required, and not on the number of items in the display (Barth et al., 2003).

3.3. Visuomotor Control

Recent work in vision science suggests that vision involves two largely separate systems: a relatively slow system concerned with the conscious perception of objects, and a faster nonattentive (and nonconscious) system concerned with the on-line control of actions such as reaching, pointing, and eye movements (Milner and Goodale, 1995). As such, the possibility arises of displays designed expressly for this second system—i.e., displays that could control a visuomotor system without the involvement of visual attention or even conscious experience. Some support exists for this possibility. Visuomotor actions such as touch selection are faster and more accurate in the lower part of the visual field, where they apparently draw upon specialized representations for action (Po et al., 2004). And removing visual feedback can help a user to aim a laser pointer at a given location, an effect counterintuitive from the viewpoint of conscious perception (Po et al., 2003).

An interesting application of this would be displays that help a user move a mouse to a given location more quickly. In such a situation, there would not be any awareness of control: the user would simply “do the right thing” (Rensink, 2002b). An even more

interesting—although highly speculative—possibility is that displays could be designed to coordinate visuomotor systems (and perhaps other processes) to allow the user to carry out highly sophisticated operations without their being aware of it, much like the “auto-pilot” experience occasionally encountered by drivers, in which conscious control temporarily vanishes (see Norretranders, 1999). If this kind of control could be achieved, it would be an important step towards the development of systems that enable highly effective interaction between humans and machines, forming a system that in many ways would be a hybrid of the two, capable of drawing on the strengths of each (Clark, 2003; Rensink, 2007).

4. CONCLUSIONS

This chapter surveyed some of the recent advances in our understanding of human vision, and discussed their implications for the management of visual attention in graphic displays. Among these advances is the recognition that attention is not the “central gateway” to visual perception, but is instead just one of several quasi-independent systems, each capable of sophisticated processing even in the absence of attention. It also appears that the experience of seeing is not supported by a dense, static representation that accumulates results in a task-indifferent way, but is instead supported by a dynamic co-ordination of attention that depends on the knowledge of the observer and the particular task being done. Other kinds of processes (such as motor control) are also carried out concurrently with this, even in the absence of conscious awareness.

Such a view has several important implications for the management of visual attention. To begin, it suggests that the ability to send attention to a particular item or location is not a secondary aspect of perception, but is fundamental to the creation of the picture we experience. By appropriate coercion of the control mechanisms used in everyday seeing, attention can be

managed such that it minimally intrudes upon a viewer's experience. Meanwhile, the existence of nonattentive systems—each with its own form of visual intelligence—provides additional ways of achieving high attentional efficiency, e.g., by creating perceptual structures that help with attentional allocation and engagement. Finally, the possibility also arises of using nonattentive mechanisms to carry out some of the processing that would have been expected of attention.

Thus, the prospects for the effective management of attention extend beyond what would have been imagined had only our casual intuitions of visual perception been used. Among other things, the developments outlined here suggest that—if done effectively—the management of attention can enable humans and machines to seamlessly combine their strengths, resulting in systems with new levels of sophistication.

References

- Albrecht-Buehler, C., Watson, B, and Shamma, D.A. (2005). Visualizing live text streams using motion and temporal pooling. *IEEE Computer Graphics and Applications*, **25**, 52-59.
- Allport, D.A. (1993). Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years. In DE Meyer and S Kornblum (eds.), *Attention and Performance XIV*. Cambridge MA: MIT Press, pp. 183-218.
- Alvarez, G.A., and Scholl, B.J. (2005). How does attention select and track spatially extended objects? New effects of attentional concentration and amplification. *Journal of Experimental Psychology: General*, **134**, 461-476.
- Ariely, D. (2001) Seeing sets: Representation by statistical properties. *Psychological Science*, **12**, 157- 162
- Ballard, D.H., Hayhoe, M.M., Pook, P.K., and Rao, R.P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, **20**, 723-767.
- Bang, M. (2000). *Picture This: How Pictures Work*. New York: Seastar Books. pp. 62-63.
- Barth, H., Kanwisher, N., and Spelke, E.S. (2003). The construction of large number representations in adults. *Cognition*, **86**, 201-221.
- Bartram, L., and Stone, M. (2007). Whisper, don't scream: Characterizing subtle grids. *IEEE Visualization 2007, Sacramento CA*.
- Bartram, L., Ware, C., and Calvert, T. (2001). Filtering and integrating visual information with motion. *Proceedings on Information Visualization*, pp. 66-79
- Block, B.A. (2001). *The Visual Story: Seeing the Structure of Film, TV and New Media*. Boston MA: Focal Press; p. 132.
- Blok, C.A. (2006) Interactive animation to visually explore time series of satellite imagery. In S.

- Bres and R. Laurini, (eds.), *Visual Information and Information Systems*. Berlin: Springer, pp. 71-82.
- Bruce, N., and Tsotsos, J.K. (2006). Saliency based on information maximization. *Advances in Neural Information Processing Systems*, **18**, 155-162.
- Burton, A.M., Bindemann, M., Langton, S.R., Schweinberger, S.R. and Jenkins, R. (2009). Gaze perception requires focused attention: Evidence from an interference task. *Journal of Experimental Psychology: Human Perception and Performance*, **35**, 108-118.
- Card, S.K., Mackinlay, J.D., and Shneiderman, B. (1999). Information visualization. In Card, S.K., Mackinlay, J.D., and Shneiderman, B. (eds.), *Readings in Information Visualization: Using Vision to Think*. San Francisco: Morgan Kaufman, pp. 1-34.
- Chong, S.C., and Treisman, A. (2003). Representation of statistical properties. *Vision Research*, **43**, 393-404
- Chun, M. M., and Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, **36**, 28-71.
- Clark, A.J. (2003). *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Cambridge MA: MIT Press.
- DeCarlo, D., and Santella, A. (2002). Stylization and abstraction of photographs. *SIGGRAPH 2002*, pp. 769-776.
- Egeth, H.E., and Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, **48**, 269-97.
- Elazary, L. and Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision*, **8**, 1-15.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, **37**, 85-104.

- Few, S.C. (2004). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Oakland, CA: Analytics Press, pp. 92-130.
- Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (ed.), *Perception* (Vancouver Studies in Cognitive Science) (Vol. 5). New York: Oxford University Press, pp. 89-110.
- Halper, N., Mellin, M., Herrmann, C.S., Linneweber, V., and Strothotte, T. (2003). Towards an understanding of the psychology of non-photorealistic rendering. *Proc. Workshop Computational Visualistics, Media Informatics and Virtual Communities* (April 4-5, 2003), Wiesbaden. Deutscher Universitäts-Verlag, pp. 67-78
- Hayhoe, M.M., Bensinger, D.G., and Ballard, D.H. (1998). Task constraints in visual working memory. *Vision Research*, **38**, 125–137.
- Healey, C.G. (1996). Choosing effective colors for data visualization, *Proceedings of the 7th IEEE conference on Visualization '96*. San Francisco, pp.263-270
- Healey, C.G., Booth, K.S., and Enns, J.T. (1996). High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interaction*, **3**, 107-135.
- Hochberg, J.E. (1978). *Perception (2nd ed.)* Englewood Cliffs, NJ: Prentice-Hall, pp. 158-211.
- Intriligator, J., and Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, **43**, 171–216.
- Itti L. (2005). Models of Bottom-Up Attention and Saliency. In L. Itti, G. Rees, J. K. Tsotsos (eds.), *Neurobiology of Attention*. San Diego: Elsevier, pp. 576-582
- Itti, L, Rees, G., and Tsotsos, J.K. (2005). *Neurobiology of Attention*. San Diego: Elsevier.

- Kamba, T., Elson, S.A., Harpold, T., Stamper, T., and Sukaviriya, P. (1996). Using small screen space more efficiently. *CHI 96*, pp. 383-390.
- Kosara, R., Miksch, S., and Hauser, H. (2002). Focus and context taken literally, *IEEE Computer Graphics and its Applications*, **22**, 22-29.
- Khan, A., Matejka, J. Fitzmaurice, G., and Kurtenbach, G. (2005). Spotlight: Directing users' attention on large displays. *Proceedings of ACM CHI 2005*, pp. 791-798.
- Kuhn, G., Amlani, A.A., and Rensink, R.A. (2008). Towards a science of magic. *Trends in Cognitive Sciences*, **12**, 349-354.
- Laurel, B. (1993). *Computers as Theatre*. Reading, MA: Addison-Wesley.
- MacEachren, A.M. (1995). *How Maps Work: Representation, Visualization, and Design*. New York: Guilford Press, pp. 51-149.
- Mack, A., and Rock, I. (1998). *Inattentional Blindness*. Cambridge, MA: MIT Press
- Marcel, A.J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, **15**, 197-237.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.
- Massironi, M. (2002). *The Psychology of Graphic Images: Seeing Drawing, Communicating*. Mahwah NJ: Erlbaum.
- McCrickard, D.S., and Chewar, C.M. Attuning notification design to user goals and attention costs. *Communications of the ACM*, **46**, 67-72.
- Milner, A.D., and Goodale, M.A. (1995). *The Visual Brain in Action*. Oxford: University Press.
- Moore, C.M., and Egeth, H. (1997). Perception without attention: Evidence of grouping under

- conditions of inattention. *Journal of Experimental Psychology: Human Perception and Performance*, **23**, 339-352.
- Mullet, K. and Sano, D. (1995). *Designing Visual Interfaces*. Englewood Cliffs, NJ: Prentice-Hall, pp. 51-87.
- Nakayama, K., He, Z. J., and Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In S. M. Kosslyn and D. N. Osherson (eds.), *Visual Cognition* (2nd ed.). Cambridge, MA: MIT Press, pp. 1-70.
- New, J., Cosmides, L., and Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, **104**, 16598-16603.
- Norretranders, T. (1999). *The User Illusion: Cutting Consciousness Down to Size*. New York: Penguin Books.
- Oliva, A. (2005). Gist of a scene. In L. Itti, G. Rees, J.K. Tsotsos (eds.), *Neurobiology of Attention*. San Diego: Elsevier, pp 251-256.
- Palmer, S.E. (1999). *Vision Science: Photons to Phenomenology*. Cambridge MA: MIT Press
- Parasuraman, R. (ed.) (2000). *The Attentive Brain*. Cambridge, MA: MIT Press.
- Pashler, H. (1998). *Attention*. San Diego: Psychology Press.
- Po, B.A, Fisher, B.D, and 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.
- Po, B.A, Fisher, B.D, and Booth, K.S. (2004). Mouse and touchscreen selection in the upper and lower visual fields. In *Proceedings of the SIGCHI conference on Human factors in Computing Systems*, pp. 359-366.

- Pylyshyn, Z.W. (2003). *Seeing and Visualizing: It's Not What You Think*. Cambridge, MA: MIT Press.
- Pylyshyn Z.W., and Storm, R.W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, **3**, 179-197.
- Rensink, R.A. (2000). The dynamic representation of scenes. *Visual Cognition*, **7**, 17-42.
- Rensink, R.A. (2001). Change blindness: Implications for the nature of attention. In MR Jenkin and LR Harris (eds.), *Vision and Attention*. New York: Springer, pp. 169-188.
- Rensink, R.A. (2002a). Change detection. *Annual Review of Psychology*, **53**, 245-277.
- Rensink, R.A. (2002b). Internal vs. external information in visual perception. *Proceedings of the Second International Symposium on Smart Graphics*, pp. 63-70.
- Rensink, R.A. (2003). Visual attention. In L Nadel (ed.), *Encyclopedia of Cognitive Science*, London: Nature Publishing Group, pp. 509-515.
- Rensink, R.A. (2007). The modeling and control of visual perception. In W. Gray (ed.), *Integrated Models of Cognitive Systems*. New York: Oxford University Press, pp. 132-148.
- Rensink, R.A., and Cavanagh, P. (2004). The influence of cast shadows on visual search. *Perception*, **33**, 1339-1358.
- Rensink, R.A., and Enns, J.T. (1998). Early completion of occluded objects. *Vision Research*, **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. *Psychological Science*, **8**, 368-373.
- Ristic, J., Wright, A., and Kingstone, A. (2007). Attentional control and reflexive orienting to gaze and arrow cues. *Psychonomic Bulletin & Review*, **14**, 964-969.

- Roda, C. Thomas, J. (eds.) (2006). Attention aware systems. Special issue of *Computers in Human Behaviour*, vol **22**.
- Seif El-Nasr, M. (2005). Intelligent lighting for game environments. *Journal of Game Development*, **1**, 17-50.
- Sharpe, S. (1988) *Conjurers Psychological Secrets*. Calgary AB: Hades Publications, pp. 46-83.
- Shive, J. and Francis, G. (2008). Applying models of visual search to map display design. *Int. J. Human-Computer Studies*, **66**, 67-77.
- Simons, D.J., and Chabris, C. (1999). Gorillas in our midst: sustained inattention blindness for dynamic events. *Perception*, **28**, 1059-1074.
- Solso, R.L. (1994). *Cognition and the Visual Arts*. Cambridge MA: MIT Press, pp. 149-150.
- St. John, M., Smallman, H. S., Manes, D. I., Feher, B. A., and Morrison, J. G. (2005). Heuristic automation for decluttering tactical displays. *Human Factors*, **47**, 509-525.
- Su, S.L., Durand, F., and Agrawala, M. (2005). De-emphasis of distracting image regions using texture power maps. In *Texture 2005: Proceedings of the 4th International Workshop on Texture Analysis and Synthesis*, pp. 119-124.
- Tatler, B.W. (2002). What information survives saccades in the real world? In J. Hyönä, D.P. Munoz, W. Heide, and R. Radach (eds.), *Progress in Brain Research*, **140**, 149-163.
- Theeuwes, J. and Godijn, R. (2002). Irrelevant singletons capture attention: evidence from inhibition of return. *Perception & Psychophysics*, **64**, 764-770.
- Thomas, J.J., and Cook, K.A. (2005). (eds.). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. Richland WA: National Visualization and Analytics Center.
- Thorpe, S.J., Fize, D, and Marlot, C. (1996). Speed of processing in the human visual system.

- Nature*, **381**, 520-522.
- Tognazzini, B. (1993) Principles, techniques, and ethics of stage magic and their application to human interface design. *INTERCHI 96*. New York: ACM, pp. 355-362.
- Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology*, **40A**, 201-237.
- Tufte, E.R. (1990). *Envisioning Information*. Cheshire CT: Graphics Press, pp. 53-65.
- Tufte, E.R. (2001). *The Visual Display of Quantitative Information*. 2nd edition. Cheshire CT: Graphics Press.
- Vertegaal, R. (ed.). (2003). Attentive user interfaces. Special issue of *Communications of the ACM*, vol. **46**.
- Vishwanath, D. and Kowler, E. (2003). Localization of shapes: Eye movements and perception compared. *Vision Research*, **43**, 1637-1653.
- von Mühlenen, A., and Lleras, A. (2007). No-onset looming motion guides spatial attention. *Journal of Experimental Psychology: Human Perception & Performance*, **33**, 1297-1310.
- Ward, P. (1996). *Picture Composition for Film and Television*. Oxford: Butterworth-Heinemann, pp. 45-68.
- Ware, C. (2004). *Information Visualization: Perception for Design (2nd edition)*. San Francisco: Morgan Kaufman.
- Ware, C. (2008). *Visual Thinking for Design*. San Francisco: Morgan Kaufman.
- Wickens, C.D. and McCarley, J.S. (2008). *Applied Attention Theory*. Boca Raton, FL: CRC Press.
- Wolfe J.M. (1999). Inattentional amnesia. In V. Coltheart (ed.), *Fleeting Memories*. Cambridge, MA: MIT Press, pp. 71–94.

Wolfe, J.M. (2000). Visual attention. In K.K De Valois. (ed.), *Seeing*. 2nd ed. San Diego: Academic Press, pp. 335-386.

Wolfe J.M. (2005). How might the rules that govern visual search constrain the design of visual displays? *Society for Information Display*. Boston MA.

Wolfe, J.M., and DiMase, J.S. (2003). Do intersections serve as basic features in visual search? *Perception*, **32**, 645-656.

Acknowledgements

Many thanks to Minjung Kim and Lonnie Hastings for their comments on earlier drafts of this chapter. Much of the cited work done by the author was carried out at Cambridge Basic Research (CBR), a laboratory of Nissan Motor Company in Cambridge MA, USA. Support for the writing of this chapter was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), and The Boeing Company.

Footnotes

1. As used here, “graphic” denotes displays that employ graphic elements, without reference to any viewer. In contrast, “visual” displays involve extensive use of visual intelligence for their operation, and thus depend strongly on the involvement of a viewer. In a way, the distinction is primarily one of emphasis: means (graphics) vs. ends (vision). For the issues discussed here, this distinction is not a critical one.
2. As used here, “control” refers to the management of attention in everyday seeing (including relevant mechanisms), while “coercion” refers to the management of these control mechanisms via particular contents of the display (essentially a second-order form of control). Control is done internally; coercion is done by external means (i.e., via the display), ideally to improve the management of attention for the task at hand.
3. “Efficiency” is sometimes used as a synonym for search rate (e.g., Wolfe, 2000). In this chapter, however, it is used in its more basic comparative sense—a measure of how close performance corresponds to that of an optimal system (where attentional mechanisms are used in the best way possible). Such attentional efficiency can therefore be viewed—at least in principle—as a percentage that can range between 0 and 100.