Attention, Consciousness, and Data Display

Ronald A. Rensink
Departments of Psychology and Computer Science, University of British Columbia

Abstract

Recent advances in our understanding of visual perception have shown it to be a far more complex and counterintuitive process than previously believed. Several important consequences follow from this. First, the design of an effective statistical graphics system is unlikely to succeed based on intuition alone; instead, it must rely on a more sophisticated, systematic approach. The basic elements of such an approach are outlined here, along with several design principles. An overview is then given of recent advances in our understanding of visual perception, including rapid perception, visual attention, and scene perception. It then is argued that the mechanisms involved can be successfully harnessed to allow data to be displayed more effectively than at present. Several directions of development are discussed, including effective use of visual attention, the display of dynamic information, and the effective use of nonattentional and nonconscious perceptual systems.

Keywords: Perception, Attention, Vision, Change Blindness, Statistical Graphics, Data Visualization, Information Visualization, Visual Design.

1. Introduction

Graphics has long been used for the display of data (Massironi, 2002; Tufte, 2001). But how can we be sure that a given system is *effective*—i.e., that it enables performance that is rapid, accurate, and accurate? Can we do better than just “look and see”, or use *post hoc* measures? More adventurously, could we determine constraints on what might be optimally effective, and incorporate these into a design ahead of time? And could this be done in a way that helps guide the design not only of static displays, but of dynamic displays that include animations of various kinds?

It might be argued that issues such as these are not that important, since designers over the years have explored (largely by brute-force trial and error) most of the space of possible designs, and now have a reasonably good idea of what the important factors are. It might also be argued that we as observers have extensive (and perhaps even privileged) experience with the operation of the human visual system, and so can easily determine the issues that would be relevant in any particular design.

However, it is argued here that while intuitions about design—especially those based on long experience—are useful, they are not enough. To begin with, many devices commonly used to display quantitative data (e.g., box plots, dot-dash plots, small multiples) are relatively recent, and new ones continue to appear; moreover, disagreement sometimes exists as to which ones are effective (Cleveland, 1994; Kosslyn, 1994; Tufte, 2001). This is difficult to reconcile with a design space that is well understood. In addition, there are few intuitions as to how information should be displayed dynamically, not to mention how it might be handled in an interactive way (Ware, 2004).

This failure of intuition also applies to our understanding of how we see. Although we have a strong impression that we experience a complete “picture” of our surroundings at all times, recent work has shown that the visual system cannot operate this way, and instead we see via a dynamic “just in time” representation that represents only a few coherent objects at any moment (e.g. Rensink, 2000). In addition, evidence is increasing that conscious perception is only one aspect of how we see, and that considerable visual intelligence exists in processes that operate without any conscious involvement at all (e.g. Milner & Goodale, 1995). As such, designs based on simple visibility or conscious introspection alone are likely to inadequate for many if not most purposes.

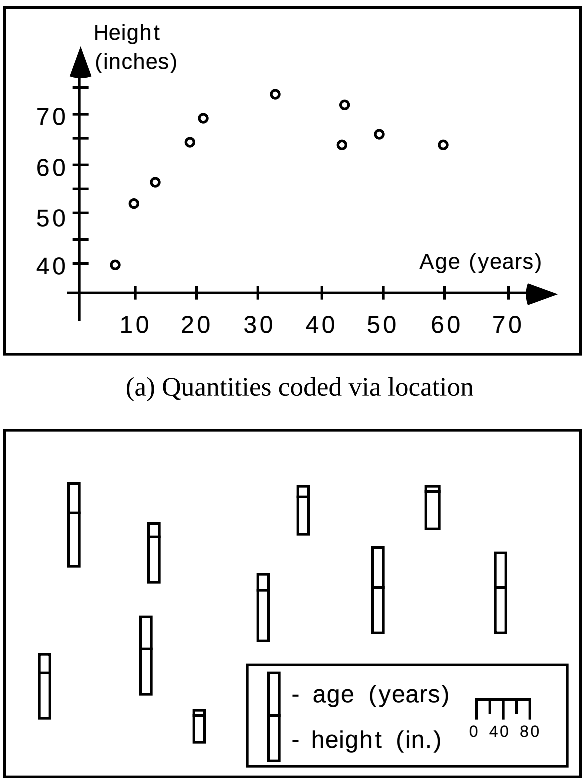
Although these considerations limit the extent to which an effective visualization system can be designed in a casual manner, they also open up the possibility of systems that are more effective than current ones. However, the realization of these will depend on design practices that take into account the nature of the mechanisms underlying human vision.

This paper discusses three important issues in the development of such practices. The first is a general framework for understanding the design of graphical representations, and developing constraints to guide it. Examples are given of explicit constraints based on current practices. Second, several recent developments in our understanding of visual perception are sketched, with emphasis placed on those aspects that are most counterintuitive. Finally, a brief discussion is given of ways in which constraints based on these new developments can be used to guide the design of new, more effective forms of data display.

2. The Design of Graphical Representations

Statistical graphics might be usefully characterized as *the transformation of statistics problems into graphical form, so as to take advantage of the visual intelligence of the human observer.* As such, the effectiveness of a given graphical representation will depend on the extent to which it allows appropriate processing mechanisms to engaged.

Consider for example a simple dataset containing the age and height of a set of people. When age and height are represented via position (Figure 1a), several trends—such as height increasing with age up to age 20—are immediately apparent. In contrast, when age and height are represented via length (Figure 1b), there is a marked reduction in the speed and ease of perceiving these relations. Trends can no longer be seen at a glance, even though the same set of data is involved in both kinds of representation.



(a) Quantities coded via location

(b) Quantities coded via length

Figure 1. Different kinds of graphical representation. (a) Age coded by horizontal position, height via by vertical position; size of the graphical elements is irrelevant. Using this representation, the relation between age and height is immediately apparent. (b) Age coded by length above interior line, height by length below; position of the graphical elements is irrelevant. Here, the relation between the two quantities is no longer obvious, even though the data represented in (a) and (b) is the same.

These two kinds of representation clearly have different levels of effectiveness. But *why*? Why are some graphical representations effective, and others not? Ideally, the explanation for this could be done in a way that would not only apply to this example, but would be general enough to be extended to all aspects of perception, and to all aspects of data visualization.

2.1 Marr's Framework

An approach that has had great success in machine and human vision is the *computational framework* of David Marr (Marr, 1982). Here, visual perception is considered to be an information-processing problem, with the explanation of any particular process involving three distinct levels of explanation:

i) *Functional* (or *computational*) This concerns the nature of the task, i.e., how the process relates input and output. For example, a functional explanation of a cash register would explain why addition—rather than multiplication, say—is the operation being carried out. The explanation for this is entirely at the functional level; it does not need to consider the mechanisms involved.

ii) *Representational (*or *algorithmic).* Analysis here concerns the nature of the representations used. In the case of the cash register, explanation at this level would account for the use of a base 10 representation, rather than e.g., base 2, base 16, or Roman numerals.

iii) *Implementational.* Here, explanation focuses on the hardware used. For the cash register example, an explanation at this level would account for the particular electronic circuits used, showing for example, that they can be used to encode numerals (of whatever base), as well as being economically viable and physically robust.

This framework has proven remarkably successful, allowing statements to be made about the general nature of complex information-processing problems (and their solutions), while at the same time clearly separating these from more particular concerns, such as the kinds of representations used by humans or by machines. In addition, such constraints have also provided useful guidelines for the design of more effective machine vision systems (Marr, 1982).

2.2 A Framework for Statistical Graphics

Given the success of the computational framework for explaining visual processes in existing (biological) systems and for guiding the design of new (artificial) ones, it is worth considering whether it might be adapted to the issues considered here. Statistical graphics (and more generally, data visualization) is at heart an information-processing problem, and so is potentially susceptible to this kind of analysis. The question is then whether the three levels might also be applied to this situation in a natural way. This appears to be possible, *provided that the observer plus the graphics system is regarded as a single information-processing system*. In this view, information is taken from an external source of data, processed—simpler operations done by the graphical system (whether static or dynamic), more complex by the visual system—and the output then determined (Figure 2). In essence, the human visual system is the "backend" of a combined system, with visual intelligence providing the means by which sophisticated processing can be carried out.

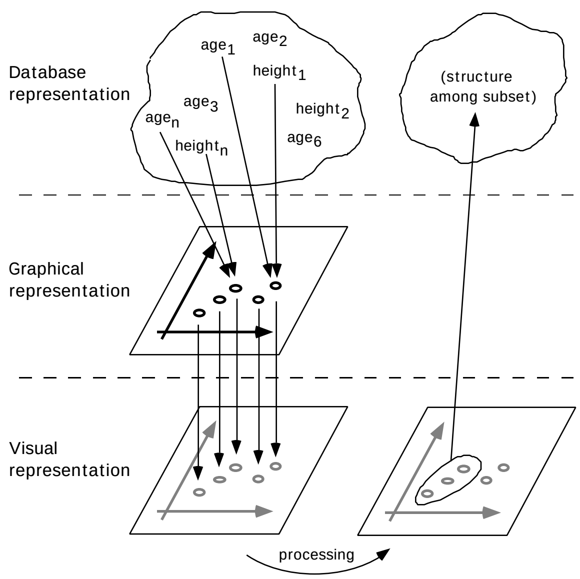


Figure 2. Data visualization. Information is taken from an external source and processed to find interesting structure. The transformation of data needed for this is done in two stages: a transformation of the relevant dimensions of the data (in whatever representation used by the database or repository) to a graphical representation on an external device (e.g., a computer display or a printed diagram), followed by a transformation of the graphical representation to a visual one in the human observer. The visual system then processes the visual representation (via innate mechanisms) to find the required structure. The form of the first transformation is subject only to general constraints on graphical elements (e.g., the laws of geometry), along with constraints due to the complexity considerations (for systems where real-time transformation is required). Transformation at the second stage is carried out by the human visual system and so is largely fixed, although it might be affected by high-level expectation. The two transformations involve successive processing stages. As such, the design of the first transformation must be such that that when coupled with the second, the operation of the combined system is effective, i.e., is as rapid, accurate, and effortless as possible.

The combined observer+graphics system can therefore be regarded as a single system that "views" the data in an abstract data space. Done effectively, this is essentially a way of extending human visual perception into completely new realms (see e.g., Clark, 2003).

Given this characterization, Marr's three levels of explanation apply as follows:

i) *Functional*. This concerns the task of the combined system: given a particular data set, detect the interesting structure it contains. Such structure could be correlations of various kinds, outliers, etc.; the particular choice will depend on the task. For instance, in the case of the age-height example, the task is to find interesting correlations between the two data dimensions.

ii) *Representational*. Here, analysis concerns the representations used. For a visualization system, two such representations are central. The first is *graphical*, i.e., the representation of the data in pictorial form on an external device. The second is *visual*, i.e., the representation in the human observer. Graphical and visual representations support two successive stages of processing, with the graphical being the one most easily controlled.

iii) *Implementation*. This concerns issues of hardware in both the system and the observer. (Although the flow of information is from external source to human, processing occurs in both "components" concurrently.) In the interests of brevity, this level will not be discussed further here.

2.3 Nature of the Design Task

According to the framework above, the detection of patterns via data visualization is a form of perception, with the design of effective displays a problem at the representational level. In particular, the design problem is to find a graphical representation such that its elements invoke visual elements and operations that compute the requested structure effectively.

To flesh this out a bit, consider again the age-height example of Figure 1a. Visualization here proceeds in two stages. The first maps the data dimensions of age and height onto the graphical dimensions of horizontal and vertical position. The original problem of finding correlations among abstract data now becomes one of finding correlations among the positions of graphical elements; this can be done via grouping on the basis of spatial location, with the overall extent and orientation of the group reflecting the nature of the correlation. Once a geometrical correlation has been found (by whatever means), it can then be mapped back to the original domain, and interpreted as a correlation in the abstract data space. Looked at this way, data visualization embodies a type of *similarity transform*—a technique commonly used in problem solving that transforms the initial representation of a problem into one that enables the solution to be easily found (Melzak, 1983).

The second stage maps the graphical representation to a visual one. The visual representations formed correspond to the graphical elements, with locations in the perceptual domain being similar to those in the graphical one. For the representation of Figure 1a, the visual system applies innate mechanisms to easily group these elements on the basis of their perceived location; the resulting groups then provide the basis for perceived structure. Meanwhile, the ineffective representation of Figure 1b also maps age and height onto a distinct graphical property (length). But there is no visual mechanism that can group perceptual elements on the basis of length above or below an arbitrary mark. As such the automatic determination of correlation fails for this kind of representation.

It is important to note that the critical factor here is not the graphical representation *per se*. At the graphical level, grouping is just the linking of related elements, and could be done on the basis of any distinct property; a simple computer program could, for instance, take the graphical representation shown in Figure 1b and extract the correlations from it. But there is no guarantee that a given graphical operation can always be carried out by visual operations that are effective. Indeed, it might even be that some graphical operations cannot be carried out by visual process at all. Thus, the critical factor of effective display is *the ability of the graphical representation to invoke the appropriate visual elements and operations for the task at hand*.

2.4 Operations Inventory

A useful way of taking human perception into account when designing a system for data display would be via an *operations inventory*, a catalog that lists the set of all possible graphical operations, along with the most suitable graphical representations for each. The time, effort, and accuracy associated with each representation would be listed; no other aspects of the visual system would need to be considered. (Although constraints such as operations that cannot be done concurrently could also be listed.) The design problem would then be one of using this inventory to find the representation that enables the least-expensive sequence of geometric operations for the problem at hand. Finding this sequence would still be a problem, but it now would be an objective, well-defined one. And the result would be a representation guaranteed to enable performance that is as rapid, accurate, and effortless as possible.

2.5 Representational Principles

An operations inventory of the type sketched above does not currently exist. But a start on perceptually-informed design can be made via *representational principles*, which constrain the kinds of mappings between data values and graphical representations, so as to avoid some potential problems right from the start. Such principles can be derived from various kinds of considerations. Among the more basic are:

i) *Invertibility*. The graphical representation must enable a 1:1 mapping between data values and visual representations. If different data values mapped to the same representation, information would be lost; moreover, results could be ambiguous. Conversely, if different visual structures mapped to the same data value, different visual processes would be involved, causing performance to degrade. (This principle implies a 1:1 mapping between data and graphical representations. But this does not necessarily guarantee a 1:1 mapping between data and visual representations.)

ii) *Distinctness.* Values that need to be distinguished for the task must map onto distinct visual representations. Otherwise, important information will be lost. This principle applies in two ways: to different data dimensions, and to different values along a single dimension.

iii) *Uniformity*. Values along a single data dimension must map onto a single visual dimension. If different visual dimensions were used, different processes would be needed. Information and time would then be lost when results are combined.

iv) *Perceptual ordering.* Data values that are ordered must map to a perceptual property that is ordered in the same way (at least over the range involved). Otherwise, the ability of the visual system to use perceptual order cannot be employed; indeed, if the visual system operates against its natural ordering, performance might degrade substantially.

v) *Minimal variation.* Graphical properties that do not encode data should not vary. If such variation were to exist, it would create a source of noise that could only cause performance to slow down and become less accurate.

These principles are not entirely new—they have formed the basis (often unstated) for much of good design. For example, the principle of minimal variation is obeyed in the effective representation of Figure 1a, which uses circles of the same size and shape for each data point; in contrast, it does not help the ineffective representation of Figure 1b that its data points have irrelevant variation in their positions. And the principle of perceptual ordering, for example is often explicitly recommended for effective display (e.g. Ware, 2004), while the principle of distinctness is similar to the strategy of the smallest effective difference (Tufte, 1997). What is proposed here is that *all* such principles should be made explicit, and that they be derived in a more systematic way based on the nature of task and the nature of human perception.

3. Visual Perception

The effectiveness of a visualization system depends critically on the extent to which it enables appropriate perceptual processes to be engaged for the task at hand. Although representational principles help, they are insufficient to guarantee good design; moreover, they still require at least some understanding of visual perception. As such, it is important to obtain some knowledge about the processes involved, especially those for which our intuitions are not always accurate.

In what follows, discussion will focus on three particularly important sets of processes: (i) that that act rapidly, prior to visual attention, (ii) those that are slower and require attention, and (iii) mechanisms that co-ordinate these to create scene representations. Only a rough sketch will be presented here; for further details, see e.g., Palmer (1999).

3.1 Rapid Vision

Effective visualization relies heavily on processes that are *rapid*, i.e., act within a few hundred milliseconds (Rensink & Enns, 1995). These require little or no attention, and are not greatly influenced by observer expectations. They are also carried out automatically, without any conscious awareness. As such, they provide visual intelligence that is effectively "free"—they do not usually interfere with other tasks, and only their output impacts consciousness.

Since rapid processes do not require attention or consciousness, it was long believed they were limited to simple operations, such as perceiving colors or two-dimensional orientation. However, recent work shows that they can do considerably more than that.

3.1.1 Visual organization

When light first enters the eyes, it is processed by a distinct stage of *early vision* (Marr, 1982). Processes here are both rapid and *low-level*, i.e., act in parallel throughout the visual field. They provide an initial organization of the input, forming "primitives" upon which all subsequent visual processing is based (Beck, 1982; Julesz, 1984). These primitives generally involve simple geometric and radiometric properties such as orientation, lightness, and motion (Palmer, 1999).

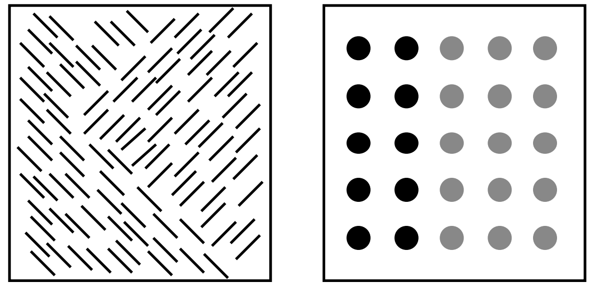


Figure 3. Visual organization. (a) Texture segmentation. Regions are formed on the basis of textures differing in a primitive property (here, orientation). (b) Grouping. Groups are formed on the basis of Gestalt principles, such as similarity (here, similarity in lightness).

Among other things, early vision provides immediate and effortless *texture segmentation*, which forms distinct spatial regions for textures differing in these properties (Figure 3a). The perceptual elements (as opposed to properties) underlying this are believed to support most aspects of texture perception, and are often referred to as *textons* (Julesz, 1984).

A related process is *grouping*, where items are rapidly organized into perceptual units on the basis of their properties (Figure 3b). Grouping is governed by *Gestalt principles* such as similarity, proximity, and good continuation (which includes alignment).

Most combinations of primitive properties (e.g., color and orientation) act independently; these are referred to as *separable* properties (Ware, 2004). A few pairs (e.g. hue and saturation, height and width) are such that performance cannot be based on just one component, but is affected by both; these are referred to as *integral* properties (Palmer, 1999)

3.1.2 Rapid detection

Visual *features* are defined as properties that "pop out", i.e., that can be rapidly detected if unique to a display, no matter how many items are present (Fig. 4a). They are not the same as the properties that govern texture segmentation (Wolfe, 1992), and thus appear to reflect a different set of visual processes.

The list of features includes color, motion, contrast, and orientation (e.g. Treisman, 1985). However, there is also evidence for considerable "visual intelligence". For example, a connected group of lines can pop out based on its overall length but not on the length of any component (Rensink & Enns, 1995). And items separated via occlusion can be linked (Fig 4b). Thus, elements of this stage are sometimes characterized as *proto-objects* (i.e., precursors to objects) rather than simple visual primitives (Rensink & Enns, 1995).

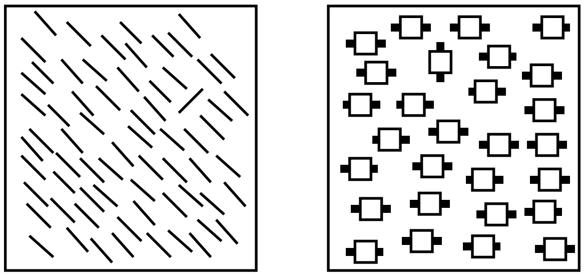


Figure 4. Rapid Detection. (a) Pop-out of items containing a unique simple feature (here, orientation). Such items are immediately detected by the visual system. (b) Features based on complex structures (proto-objects). Here, detection is based on the unique orientation of the occluded item, formed by rapid completion of the black squares.

Scene-based properties can also be encoded at this level. For example, rapid detection can be influenced by such things as three-dimensional orientation, direction of lighting, surface convexity/concavity, and shadow formation. Such estimates are formed on the basis of "quick and dirty" assumptions that are true most of the time (Rensink & Cavanagh, 2004). This supports the proposal that early vision produces a viewer-centered description of the world in which scene properties are represented in a fragmented fashion (Marr, 1982).

3.1.3 Rapid statistical estimation

Another form of rapid visual intelligence—one particularly important for statistical graphics—is the ability to form *statistical estimates* of sets of briefly-presented items. For example, observers can match the mean size of a set of disks to an individual disk as accurately as they can match the sizes of two individual disks (Ariely, 2001); accurate estimates can be formed using exposures of as little as 50 milliseconds (Chong & Treisman, 2003). This ability may also extend to other statistical measures, such as range or variance, although this has not yet been confirmed.

3.1.4 Scene composition

Rapid processing can also extract several aspects of scene composition. One of these is the abstract meaning (or *gist)* of the scene—e.g., whether it is a city, airport, or farm. This can be determined within 100 milliseconds, a time insufficient for attending to more than a few items. Other aspects of scene composition, such as how open or crowded it is, can also be obtained rapidly (see Oliva, 2005). Gist is likely determined on the basis of simple measures such as the distribution of line orientations or colors in the image. An interesting conjecture is that it may rely on the rapid statistical estimates described above.

3.2 Visual Attention

A critical factor in the conscious perception of a display is visual attention. This can be defined as the selective processing of information; it operates at relatively slow speeds—from 50-300 milliseconds per operation (see, e.g., Rensink, 2003).

Visual attention is not well understood. In fact, it may involve a set of unrelated mechanisms, having in common only the fact that they are each of limited capacity. As such, it is perhaps best to examine it via the functions it is believed to be necessary for.

3.2.1. Selective integration

An important type of attention is *selective integration* —the binding of selected parts or properties into a more complex structure. For example, it is difficult to detect a single L-shaped item among a set of T-shaped items; difficulties are also experienced for unique combinations of orientation and color, or of most other features. This has been explained by *feature integration theory*, which asserts that attention acts via a spotlight that integrates the features at each location into an *object file*, at a rate of about 50 ms per item (Treisman, 1988), If a graphical element contains a unique feature, it will be rapidly detected (Section 3.1.2); otherwise attention must be sent around the display on an item-by-item basis.

3.2.2. Change detection

In Figure 5, an original image A alternates with a modified image A', with brief blank fields between successive images. Observers have great difficulty noticing most changes under these conditions, even when the changes are large, repeatedly made, and the observer knows that they will occur. Such *change blindness* can exist for long stretches of time—up to 50 seconds in some cases (Rensink et al., 1997).

Figure 5. Change blindness. Original image A (statue with wall in background) and modified image A' (statue with wall gone) are displayed in the order A, A', A, A',…. with gray fields placed between successive images. Observers typically require several seconds to see such a change, even though it is large.

Change blindness can be accounted for by the hypothesis that attention is needed to consciously see change. (This essentially extends selective integration into the time domain.) A change will be difficult to see whenever 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. This explains why change blindness can also occur for a change made during an eye movement, eyeblink, or occlusion by some other object (Rensink, 2000).

This perspective forms the basis of *coherence theory* (Figure 6). Prior to attention, proto-objects are continually formed in parallel across the visual field. Attention selects a few of these, and enters them into a *coherence field,* a circuit between the attended items and a higher-level *nexus*. This yields a representation with spatio-temporal coherence. Attention is released by breaking this circuit, with the object dissolving back into its constituents. There is little or no "after-effect" of having been attended. (Also see Wolfe, 1999.)

When an item is attended, it is not possible to represent all of its detail in a coherence field—only a few properties can be represented at any time. If one of the properties represented is one of the aspects changing, the change will be seen; otherwise, change blindness will result, even though the item is attended.

Another important limitation arises from the pooling of information into a single nexus: attended items are not independent, but operate as a single complex. For example, the nexus cannot determine if it receives one change signal or two. Thus, observers cannot effectively see more than one change at a time, no matter how much they try (Rensink, 2001, 2002a).

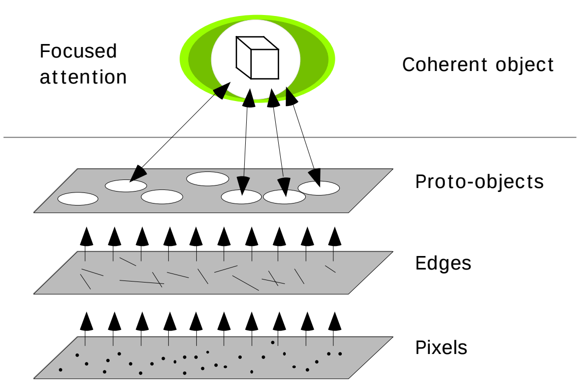


Figure 6. Coherence Theory. Early vision continually creates proto-objects across the visual field. These are volatile, lasting only a brief time. Attention "grabs" a few of these and sets up reciprocal connections between them and an attentional nexus, creating a *coherence field*. As long as the proto-objects are “held” in the field, they form an individuated object with both temporal and spatial coherence (Rensink, 2000).

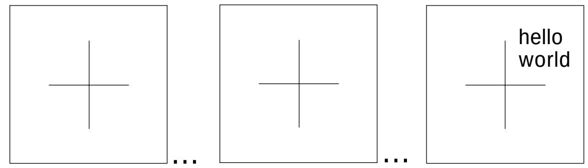


Figure 7. Inattentional blindness. Observers are shown a sequence of images, and asked to judge which line (horizontal or vertical) is longest in each. After several presentations, an image is presented that also contains an unexpected stimulus. Observers often do not see this.

3.2.3. Conscious experience

Recent studies also suggest that attention is needed for *any* conscious experience of a graphical item. For example, Mack & Rock (1998) asked observers to attend to an overlapping pair of lines (one horizontal and one vertical), and judge which was longer. After several such trials, a display was presented that contained an unexpected item (Figure 7). Observers often failed to see this item, even when it was being looked at directly. This failure is called *inattentional blindness*. Such blindness can occur even for objects that are highly visible. For example, Simons and Chabris (1999) had a person dressed in a gorilla suit unexpectedly walk across a scene; this was not seen by over half the observers tested.

Interestingly, even if an unattended object is not seen consciously, it still influences processing. For example, surrounding lines can induce a length illusion in test items, even if the lines themselves are not seen (Moore & Egeth, 1997). This reinforces the finding from studies on rapid vision that forms of visual intelligence exist that do not involve conscious awareness.

3.3 Scene Perception

The finding that attention is needed for conscious perception has several counterintuitive implications. For example, given that attention has a very limited capacity, only a few items in display could be consciously seen at any time. As such, our visual representation of a scene or display at any instant is sketchy and incomplete, containing at most only a few coherent objects. But if so, why do we not notice such limitations? Why do we believe we see all objects and all events, and these in great detail?

The answer to this involves a shift in our view of how information is integrated to perceive complete scenes. 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 dynamic rather than static terms—as *co-ordination* rather than *construction* (see Rensink, 2007).

3.3.1 Virtual representation of scenes

A possible way to account for the seeming contradiction between our impression of virtually unlimited perceptual content and recent findings of severe attentional limitations is the idea of a *virtual representation*: instead of forming a coherent, detailed representation of all the objects in our surroundings, create a coherent representation only of the item needed for the task at hand (Rensink, 2000). If attention can be managed such that a coherent representation of an item can be created whenever needed, the scene representation will appear to higher levels as if "real", i.e., as if all items 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 perception of scenes (including complete displays) relies on a relatively sparse, dynamic "just in time" system that creates most of the required representations whenever they are needed. Among other things, this has the consequence that there is no general-purpose representation used in vision: the representation in use at any moment is coupled to the task at hand, and would likely be suboptimal for other purposes. It also implies that different people literally see a scene in different ways, depending on their expectations (Rensink, 2007).

3.3.2 Triadic architecture

The successful use of virtual representation requires eye movements and attentional shifts to be made to the appropriate item at the appropriate time. But how could this be done? One possible way is via a *triadic* architecture (Figure 8):

i) an early visual system rapidly creates volatile proto-objects in parallel across the visual field.

ii) a limited-capacity attentional system forms these structures into stable object representations.

iii) a nonattentional system provides a context (or *setting*) that guides attention to the appropriate objects in the scene.

The setting system relies on properties such as gist, believed to be determined rapidly (Section 3.1.4). It also involves the spatial arrangement (or *layout*) of objects in the scene. Layout may contain the description of properties other than location; but it so, these descriptions would not be very detailed.

In this model, scene perception occurs via a continual interaction between low-level structures which automatically draw attention to interesting items, and high-level structures that guide it on the basis of existing knowledge about the objects and the scene.

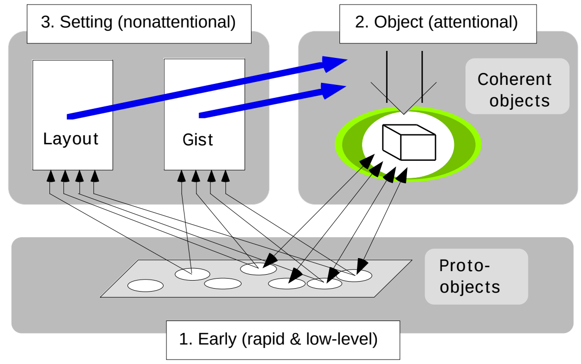


Figure 8. Triadic Architecture. (Rensink, 2000). Perception is carried out via the interaction of three systems. (1) Early-level processes create volatile proto-objects. (2) Visual attention "grabs" selected structures and forms an object with both temporal and spatial coherence. (3) Setting information (obtained via nonattentional processes) guides the allocation of visual attention.

Note that a complete representation of the scene is never constructed—there always remains only one coherent object represented at any one time. Such an approach uses representations that are stable and representations that contain large amounts of visual detail. But at no point does it use representations that are both stable *and* contain large amounts of detail.

3.3.3 Virtual Representation of items

As discussed in Section 3.2, attention monitors at most 4-5 items at a time. Moreover, only a few properties of these items can be attended. How then it possible to see even a single object in detail?

Attention is not well understood, and so a definite answer cannot be given. However, one possibility is a *virtual representation of objects*, similar to that for scenes. Only those aspects of the attended object relevant to the task at hand would be attended at any instant, with attention switching to different aspects or parts as needed. If such within-item co-ordination could be effectively managed, the result would appear to be an object with as much detail as required.

The nexus of coherence theory (section 3.2.2) lends itself well to such a proposal. It is believed to collect only a few selected (i.e., attended) properties, along with a coarse initial description of the overall shape of the item. The links to the proto-objects create a "local hierarchy", with only two levels of description (object- and part-level) in play at any time. If different structural levels could be brought into play when requested, attention could access any part of an item's structure in the image at any time, providing exactly the kind of behavior needed.

4. Implications For Design

As discussed in Section 3, recent work indicates that visual perception is not based on an internal "picture" formed by building up dense, detailed information, but is instead based on a much sparser, virtual representation that attends only to whatever information is needed at that moment. Management of this is largely guided via rapid processes having considerable visual intelligence, operating outside of conscious awareness.

Because such mechanisms are highly at odds with our impressions of how we see, it may be worth discussing how these findings could assist the design of data display systems more effective than those based on simple intuition. Some of these guidelines will take the form of representational principles of the kind discussed in Section 2; others will simply point out possible processes that might be harnessed.

4.1 Rapid Vision

4.1.1 Additional features

The properties allowing rapid detection (or "pop out") include more than simple two-dimensional features (Section 3.1.2). Thus, it is possible to use graphical dimensions beyond those employed in current systems. Care must be taken when invoking these properties, since they rely on particular assumptions, such as lighting from above. But once invoked, they should be as useful as any other dimension.

4.1.2 Failure of simple features

Rapid detection is based upon properties of relatively complex structures (proto-objects), and on not pixels in the image. Thus, a region with a distinctive size or orientation will not be available for rapid detection if it is part of such a structure. Such *pre-emption* is common in groups of connected lines, where detection is governed by the property of the group rather than by the elements (Rensink & Enns, 1995), even when these elements are noticeable on their own

4.2 Visual Attention

4.2.1 Limited information

The amount of information that can be held by attention is extremely limited: only 4-5 items in an image can be accessed at a time, with only a few properties from each item. As such, this leads to the *principle of limited information:* Only a limited amount of information—at most 2-3 properties from 4-5 items—should be conveyed dynamically. Otherwise, information will exceed attentional capacity, and pickup will be impaired.

4.2.2 Maximal Static Representation

Because attention constantly moves—both within a display as well as within objects (section 3.2)—it is best to comply with a *principle of maximal static representation*: Use static representations of data whenever possible. This not only minimizes the chances of losing information by not being at the right place when attention arrives, but also reduces distraction by minimizing irrelevant motion signals.

4.2.3 Minimal motion

Since motion signals automatically attract attention (e.g. Klein et al., 1992), irrelevant motion seriously interferes with attentional management. This is especially harmful when information is conveyed via dynamic means, since change blindness could be inadvertently created. As such, it appears useful to follow a *principle of minimal motion*: Motion should be minimized in the display whenever possible.

4.2.4 Single dynamic source

Attended information appears to be pooled into a single nexus; among other things, this creates an inability of observers to distinguish one change from two (section 3.2.2). This leads to the *principle of single dynamic source*: Use at most one dynamic information source at a time. Otherwise, if two or more such sources are attended, the observer will not be able to separate out the different signals, nor determine which one came from which source.

4.3 Scene Perception

4.3.1 Coercive displays

Given that our perception of a display depends on the careful coordination of attention (Section 3.3), it may be possible for a display to take control of attentional management and make the observer see (or not see) any given part of it. Such *attentional coercion* has long been used by magicians to achieve a variety of striking effects; by controlling attentional allocation, it may be possible to develop *coercive displays* that operate in a similar way (Rensink, 2002b). This would ensure that important items would not be missed, and would help direct attention to required locations or items. In such a situation, the observer would notice nothing unusual—the result would simply appear, as if by magic.

Acknowledgements

Nissan Motor Co. (Japan) and NSERC (Canada) supported much of the work upon which this paper is based. Thanks also to Tamara Munzner for discussions on several of these issues, and to Naomi Robbins for the opportunity to present these ideas.

References

Ariely D. (2001). Seeing sets: Representation by statistical properties. *Psychol.. Science*, 12, 157-162.

Ballard DH, Hayhoe MM, Pook PK, & Rao RP. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20, 723-767.

Beck J. (1982). Textural segmentation. In J. Beck (ed.), *Organization and representation in perception* (pp. 285-317). Hillsdale, NJ: Erlbaum.

Chong SC, & Treisman A. (2003). Representation of statistical properties. *Vision Research*, 43, 393-404

Clark A. (2003) *Natural-Born Cyborgs*. Oxford: University Press.

Cleveland WS. (1994). *The Elements of Graphing Data*. Summit, NJ: Hobart Press.

Julesz, B. (1984). A brief outline of the texton theory of human vision. *Trends in Neuroscience*, 7: 41-45.

Klein R., Kingstone A, & Pontefract A. (1992) Orienting of visual attention. In K Rayner (ed.), *Eye Movements and Visual Cognition* (pp. 46-65). New York: Springer.

Kosslyn SM. (1994). *Elements of Graph Design*. New York: Freeman.

Levin DT, & Simons DJ. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review*, 4, 501-506.

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

Marr D. (1982) *Vision*. San Francisco: Freeman.

Massironi M. (2002). *The Psychology of Graphic Images: Seeing, Drawing, Communicating*. M. Bruno, translator. Mahwah, NJ: Erlbaum

Melzak ZA. (1983). *Bypasses: A Simple Approach to Complexity*. NY: John Wiley & Sons.

Milner AD, & Goodale MA. (1995). *The Visual Brain in Action*. Oxford: Oxford Univ. Press.

Moore CM, & Egeth H. (1997). Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 339-352.

Oliva, A. (2005) Gist of the Scene. In *Neurobiology of Attention*, L. Itti, G. Rees and J. K. Tsotsos (eds.), Elsevier, San Diego, CA (pp 251-256).

Palmer SE. (1999). *Vision Science: Photons to Phenomenology*. Cambridge MA: MIT Press.

Rensink RA. (2000). The Dynamic Representation of Scenes. *Visual Cognition*, 7, 17-42.

Rensink RA. (2001). Change Blindness: Implications for the Nature of Attention. In M.R. Jenkin and L.R. Harris (eds.), *Vision and Attention* (pp. 169-188). New York: Springer.

Rensink RA (2002a). Failure to see more than one change at a time. *Journal of Vision*, 2: 245a. http://journalofvision.org/2/7/245. [VSS 2002; Sarasota, FL, USA.]

Rensink RA (2002b). Internal vs. external information in visual perception. *Proceedings of the Second International Symposium on Smart Graphics*: 63-70. [Smart Graphics 2; Hawthorne, NY, USA.]

Rensink RA (2003). Visual attention. In L Nadel (ed.), *Encyclopedia of Cognitive Science*. London: Nature Publishing Group.

Rensink RA (2007). The Modeling and Control of Visual Perception. In W. Gray (ed.), *Integrated Models of Cognitive Systems*. Oxford: University Press.

Rensink RA, & Cavanagh P. (2004). The influence of cast shadows on visual search. *Perception*, 33: 1339-1358.

Rensink RA, & Enns JT. (1995). Preemption effects in visual search: Evidence for low-level grouping. *Psychological Review*, 102: 101-130.

Rensink RA, O'Regan JK, & Clark JJ. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8, 368-373.

Simons DJ & Chabris C (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, 28, 1059-1074.

Treisman A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, 31, 156-177.

Treisman A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology*, 40A: 201-237.

Tufte ER. (1997). *Visual Explanations*. Cheshire CT: Graphics Press.

Tufte ER. (2001). *The Visual Display of Quantitative Information (2nd ed.)*. Cheshire CT: Graphics Press.

Ware C. (2004). *Information Visualization: Perception for Design (2nd ed.)*. San Francisco: Morgan Kaufman.

Wolfe JM . (1992). "Effortless" texture segmentation and "parallel" visual search are not the same thing. *Vision Research*, 32: 757-63.

Wolfe JM. (1999). Inattentional amnesia. In V. Coltheart (ed.), *Fleeting Memories.* (pp. 71-94). Cambridge, MA: MIT Press.