Internal vs. External Information in Visual Perception

Ronald A. Rensink
Departments of Computer Science and Psychology
University of British Columbia
Vancouver BC, Canada V6T 1Z4
+1-604-822-2579
rensink@cs.ubc.ca

ABSTRACT

One of the more compelling beliefs about vision is that it is based on representations that are coherent and complete, with everything in the visual field described in great detail. However, changes made during a visual disturbance are found to be difficult to see, arguing against the idea that our brains contain a detailed, picture-like representation of the scene. Instead, it is argued here that a more dynamic, "just-in-time" representation is involved, one with deep similarities to the way that users interact with external displays. It is further argued that these similarities can provide a basis for the design of intelligent display systems that can interact with humans in highly effective and novel ways.

General Terms

Design, Performance, Experimentation, Human Factors, Theory.

Keywords

Design, Vision, Attention, Change Blindness, Visual Memory.

1. INTRODUCTION

When we look at our surroundings, we generally have a strong impression of seeing all the objects in it simultaneously and in great detail. This impression is so strong that it tends to make us believe that we also *represent* all these objects simultaneously, with each object having a description that is both detailed (containing a high density of information) and coherent (all pieces combined properly). Such a description could be formed in a relatively straightforward way by accumulating information in an internal visual buffer. All subsequent visual processing can then be based on this buffer.

But does such a buffer really exist? Results from a number of recent experiments argue against such an idea. For example, if changes are made during a visual disturbance, observers often fail to notice these changes, even when these are large, anticipated, and repeatedly made. This *change blindness* [1] provides strong evidence against the idea of an internal representation that is everywhere detailed and coherent.

To explain why change blindness can be easily induced in experiments but is not evident in everyday life, it is argued that focused attention provides spatiotemporal coherence for the stable representation of one object at a time. It is further argued that the allocation of attention is coordinated to create a stable object representation whenever needed. The *virtual representation* this creates appears to higher levels as if all objects in the scene are represented in detail simultaneously.

In this view, the visual perception of a scene results from an intricate interplay between (i) internal information based on knowledge and (ii) external information about visual detail based on the incoming light. An architecture containing both attentional and nonattentional streams is proposed as a way to implement this scheme. This architecture has a number of striking similarities with the way that information is accessed on the Web, and implies that—if the display is designed correctly—the access of information from the display can be a natural extension of basic visual perception, and may even support new modes of interaction.

2. CHANGE BLINDNESS

Consider the situation shown in Figure 1. In this *flicker paradigm*, an original image A alternates with a modified image A', with brief blank fields between successive images.

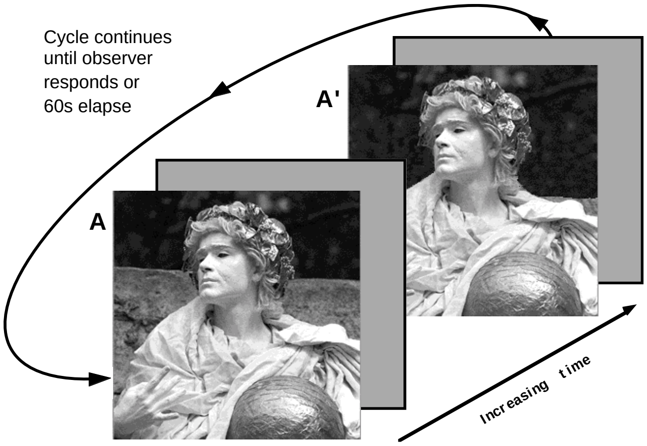

Figure 1. Design of flicker paradigm. Original image A (statue with background wall) and modified image A' (statue with wall lowered) are displayed in the order A, A', A, A ' ,... with gray fields placed between successive images [1].

Symp. on Smart Graphics, June 11-13, 2002, Hawthorne, NY, USA.
Copyright © ACM 2002 1-58113-216-6/02/07...$5.00

Interestingly, observers have great difficulty noticing most changes under these conditions, even when the changes are large, repeatedly made, and the observer knows they will occur. Such *change blindness* [1] can exist for long stretches of time—up to 50 seconds in some cases. Furthermore, it is a very general phenomenon: it can be induced in a variety of ways, such as when changes occur simultaneously with:

- image flicker
- eye movements (saccades)
- eye blinks
- occlusions by passing objects
- real-world interruptions
- movie cuts
- brief "splats" that do not cover the change
- changes made gradually

All these conditions induce a similar inability to see change. (For a comprehensive review of work on change blindness, see [2], [3].) The generality and robustness of this effect indicates that change blindness is not a phenomenon due to specialized or peripheral mechanisms, but rather, involves mechanisms central to our visual experience of the world.

3. IMPLICATIONS FOR HUMAN VISION

If a visual buffer existed, observers should be able to build up a picture of their surroundings, and then compare this picture to the current input. Given the existence of change blindness, however, it would appear that little information is being accumulated in this way. But if there is no buffer, how might human vision operate? How is it possible to see change at all?

3.1 Coherence Theory

The view of visual perception developed here is based on the proposal that *focused attention is needed to see change* [1]. Under normal circumstances, any change is accompanied by a motion signal, which attracts attention to its location (e.g.,[4]). When this local signal is swamped (via the transients associated with a saccade, flicker, eyeblink, splat, etc.), the guidance of attention is lost and change blindness is induced.

This proposal immediately runs into a challenge: According to many views of attention, it is able to "weld" visual features into relatively long-lasting representations of objects [5], and can operate at a rate of 20-40 items per second [6]. If so, why doesn't it weld all the visible items within the first few seconds of viewing and so allow the easy detection of change?

Rather than assuming that the structures formed by attention last indefinitely, it is hypothesized here that their lifetimes are actually quite brief. In particular, attention may endow a structure with a coherence lasting only as long as attention is directed to it. Developing this line of thought leads to a *coherence theory* of attention [7] (Figure 2).

Coherence theory is based on three related hypotheses:

i) Prior to focused attention, low-level *proto-objects* are continually formed rapidly and in parallel across the visual field, without attention. These "preattentive" structures can be quite complex, describing properties of the scene such as three-dimensional orientation and shadows [8]. However, they are volatile, having no real memory. Thus, they are simply *replaced* when any new stimulus appears at their location.

ii) Focused attention acts as a "hand " that selects a small number of proto-objects from this constantly-regenerating flux and stabilizes them. This is done via links that provide feedback from a single, higher-level *nexus*; when this feedback has been established, the resulting circuit is referred to as a *coherence field* [7]. This field creates a representation with a high degree of coherence over space and time. Thus, any new stimulus at that location is treated as the *change* of an existing structure rather than the appearance of a new one.

iii) After focused attention is released, the object loses its coherence and dissolves back into its constituent proto-objects. There is little or no "after-effect" of having been attended. (Also see [9].)

According to coherence theory, a change in a stimulus can be seen only if it is attended at the time the change occurs. Since only a small number of items can be attended at any time (e.g.,[10]), most items in a scene will not have a stable, detailed representation. If conditions are such that attention cannot be automatically directed to the change, the changing item is unlikely to be attended, and change blindness then follows.

The limited amount of information that can be attended at any one time also explains why observers can fail to detect changes in "attended" objects [11]. When focused attention is directed to something in the world, it will not generally be possible to represent all of its detail in a coherence field—only a few of its aspects can be represented in the nexus at any one time. If one of the aspects being represented is one of the aspects changing in the world, the change will be seen; otherwise, change blindness will again result.

This view of visual attention has several similarities with earlier proposals. In particular, the notion of a coherence field is somewhat like the proposal of an *object file* [5], which binds various properties of a spatiotemporal structure together; once set up, an object file need not be attended, so that several files can be maintained at a time. In contrast, however, only one coherence field can be set up at a time [7], [12], and it collapses as soon as attention is withdrawn.

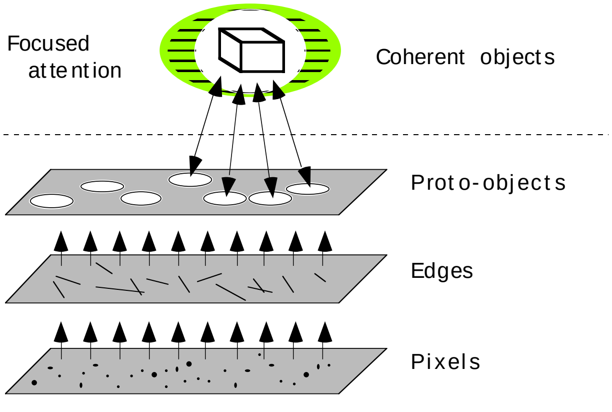

Figure 2. Coherence Theory. Early-level processes create proto-objects rapidly and in parallel across the visual field. Focused attention "grabs" these volatile proto-objects and stabilizes them. As long as the proto-objects are "held " in a coherence field, they form an individuated object with both temporal and spatial coherence [7].

3.2 Virtual Representation

3.2.1 Basics

The theory of attention proposed above has a rather counterintuitive implication: only a few items in an environment (or display) can be represented in stable form at any time. Furthermore, this representation is severely limited in the amount of information it can contain [6],[12]; as such, our representation of an environment at any given instant is at best sketchy and incomplete. But if so, why do we not notice these limitations?

To answer this, consider how objects are used in everyday life. For most tasks, only one object is in play at any time: a cup is grasped, a friend recognized, a speeding cyclist avoided. A detailed representation may be required for this "target" object, but it is not required for the others. Although there are tasks (e.g., juggling) that appear to be exceptions to this, these tasks can be handled by quickly switching back and forth, so that there is only a single target at any one time. Thus, although we may need to represent various aspects of a scene (such as the background), it appears that we may never need a detailed representation of more than one of the objects in it at any particular time.

This realization gives rise to the idea of a *virtual representation*: instead of forming a coherent, detailed representation of all the objects in our surroundings, create a coherent, detailed representation only of the object needed for the task at hand [12], [13]. If attention can be coordinated so that a coherent, detailed representation of an object can be created whenever needed, the representation of a scene will appear to higher levels as if "real", i.e., as if all objects are represented in great detail simultaneously. Such a representation will then have all the power of a real one, while using much less in the way of processing and memory resources.

3.2.2 Conditions for Successful Operation

A virtual representation can provide large savings in computational resources. Only a fraction of all visible objects need to be represented in detail, resulting in considerable savings in memory. And only a fraction of all visible objects need to be formed into coherent representations, resulting in considerable savings in processing time.

However, such savings are not possible for all problems: virtual representations reduce complexity in space by trading off for an increased complexity in time, and only certain kinds of information-processing tasks can take advantage of this. In the case of scene perception, what is required for the successful operation of a virtual representation is:

(i) only a few objects need to have a coherent representation at any moment, and

(ii) detailed information about any object must be available when requested.

The first requirement (sparseness of object representations) is easily met for most if not all visual tasks. We usually need to attend to only one object at a time, e.g., to grasp it. Tasks where several independent objects are involved can generally be handled by "time-sharing", i.e., by rapidly switching attention back and forth between the objects.

The second requirement (access on request) is also met under most conditions of normal viewing. Provided that there is a way to guide eye movements and attentional shifts to the location of the requested object, visual detail can be obtained from the stream of incoming light, and a coherent representation then formed. Consequently, a high-capacity internal memory for objects is not needed: the information is almost always available from the world itself, which—as pointed out as far back as the 1950s [3] — can effectively act as an " external memory " .

In this view, then, perception involves a partnership between the observer and their environment: rather than building up an entire internal re-creation of the incoming image, the observer simply trusts the visual world to be an external memory, i.e., providing external information whenever needed. Problems can arise when light is not available to carry the information from the object to the eyes, or when the objects themselves are somehow occluded. But these conditions also interfere with object perception itself, regardless of the memory scheme used, and so do not form a serious obstacle to the use of virtual representation.

3.3 Triadic Architecture

3.3.1 Basics

The successful use of virtual representation in human vision requires that eye movements and attentional shifts be made to the appropriate object at the appropriate time. But what directs these movements and shifts?

Several solutions are possible. One possibility consistent with what is known of human vision is a *triadic architecture* with three largely independent systems [7] (Figure 3):

(i) an early-level visual system that rapidly creates detailed, volatile proto-objects in parallel across the visual field.

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

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

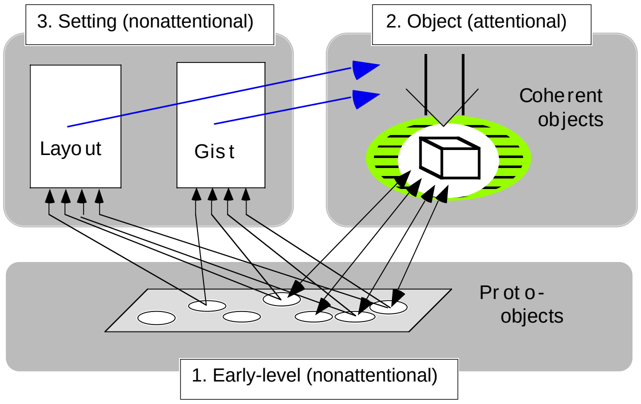

Figure 3. Triadic Architecture. Visual perception may be carried out via the interaction of three systems. (1) Early-level processes create volatile proto-objects. (2) Focused attention acts as a hand to "grab" these structures and form an object with both temporal and spatial coherence. (3) Setting information—obtained via a nonattentional stream—guides the allocation of focused attention. [7]

Note that the setting system involves at least two aspects of scene structure, both of which can be obtained largely without focused attention:

(i) The abstract meaning (or *gist*) of a scene—e.g., whether the scene is a harbor, city, or picnic. Gist can be obtained very rapidly, probably without attention [3]. It could provide a useful way to prioritize attention, directing it to the objects that are most important for the task at hand.

(ii) Perception of the spatial arrangement (or *layout*) of objects in the scene. This could rely on a nonvolatile representation of the locations of several (nondetailed) structures, which could then be used when attention is to be directed to particular objects in the scene. Note that although layout information is nonvolatile, it is not detailed, and so relatively little information is stored concerning each item.

3.3.2 Interaction of Systems

In the triadic architecture, the representation of a scene involves the dynamic interaction of three different systems. This might be carried out in the following way:

- When a scene is viewed, rapid low-level processes provide a constantly-regenerating sketch of the properties visible to the observer.
- Gist is determined by a subset of these, with subsequent processes attempting to verify the schema invoked.
- Items consistent with the schema need not be encoded in detail, since verification may involve a simple checking of expected features.
- If an unexpected structure in the image is encountered, attentional processes could form a coherent representation of its structure, attempt to determine its semantic identity, or reevaluate the gist. Meanwhile, the layout could be used to check the current interpretation, as well as help guide attention to a requested object.

According to this view, then, 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.

Furthermore, this view also contains a fundamental change of perspective: internal representations are no longer detailed structures *built up* from eye movements and attentional shifts, but rather, are sparse structures that *guide* those activities that obtain detailed information from external sources.

3.4 Nonattentional Perception

The triadic architecture outlined in Section 3.3 is based on a view of attention quite different from the traditional one: rather than being the "main gateway " of all visual perception, attention is just one of several concurrent streams, namely the stream concerned with the conscious perception of coherent objects. The other streams do not rely on attention, and so can operate independently of it

Although relatively little is known about these nonattentional streams, all appear to be involved with perceptual processes that operate without the conscious awareness of the observer; indeed, it may be that perception without attention is exactly perception without awareness [14].

Examples of such nonattentional streams are the systems that underlie motor actions such as reaching, grasping, and eye movement. These systems have neural circuitry that differs from that underlying attentional (conscious) perception; indeed, under some conditions they can even behave in ways counter to what is consciously experienced [15].

Another example is subliminal (or *implicit*) perception. When a stimulus is masked almost immediately after being presented, it can cause priming (i.e., an increased sensitivity to the subsequent occurrence of the same stimulus), even though the observer was unaware of it [16].

Another interesting phenomenon in this regard is *mindsight*, where observers watching a flicker display can "sense " that a change is occurring, even though they do not have a visual experience of it [3]. Although this phenomenon is poorly understood, it appears likely that nonattentional mechanisms are involved.

4. IMPLICATIONS FOR DISPLAYS

The view put forward here suggests that the representations used in human vision can either be detailed (proto-objects) or stable (coherence fields), but never both detailed *and* stable. In other words, visual perception is never based on an internal "picture" formed by building up detailed information over time, but instead is based on a dynamic virtual representation that stabilizes only whatever information is needed at that moment. As such, there is no general-purpose representation used in vision: the representation in play at any moment is coupled to the task at hand, and would likely be suboptimal for other purposes.

It is worth pointing out that virtual representation is a special case of deictic (or indexical) representation. Such representations do not *construct a copy* of the world or of their neighbors—rather, they *coordinate* the actions of the various systems involved [13], [17]. It is important to point out that this is not limited to use of external *memory*—for example, the world can also be used as an external *processor* [18]. Thus, provided that coordination is handled well, the use of artifacts for perceptual and cognitive amplification can become a completely natural extension of human experience. Furthermore, the possibility also arises of visual displays that can support new modes of perception and interaction.

A few examples of these possibilities are sketched below. These will hopefully give some indication of the kinds of opportunities that can result from taking this viewpoint into account when designing visual displays.

4.1 Attentional Pickup of Information

4.1.1 Basics

One of the main points of the view of human vision sketched in Section 3 is that it involves a constant interaction with the outside world. The amount of information that can be held by attention (information which appears to underlie our conscious visual experience [3], [14]) is extremely limited: only 4-5 items in an image can be accessed at a time, with only a few properties from each item. Creating the conscious impression of a rich, coherent display requires the careful coordination of attention with the task at hand, so that external information can always be accessed when needed.

If a display is to be designed so that the observer can interact with it optimally, it is essential to understand how attention operates. Unfortunately, there is still much about attention that is not known, and there exist several different—and somewhat incompatible—theories that attempt to describe it (e.g. [5], [10]). However, a reasonable amount of convergence does exist among most of these, so that choosing any single framework should provide a reasonable first approximation for most purposes.

In what follows, the framework used will be coherence theory. Coherence theory has the advantage of being compatible with many aspects of other current theories [3], while also being able to account for the striking effects encountered in change-blindness studies. And as a result of the wide variety of effects that have been found [12], coherence theory has a relatively high degree of articulation, both in the mechanisms proposed, and the predictions that it can make. Again, it is unlikely that this theory will be the final story, but it should provide a good starting point.

As discussed in Section 3.1, coherence theory posits that attention acts via a coherence field that links 4-5 proto-objects to a single nexus. The nexus collects a few selected (i.e., attended) properties, along with a coarse description of the overall shape of the item. The information in the nexus essentially describes the whole object perceived at any given moment, with the links providing connections to its parts. As such, this representation is a "local hierarchy", with only two levels of description (object- and part-level). Such a hierarchy is an extremely useful device, and is a natural way to represent objects [19].

Importantly, a proto-object can be attentionally subdivided and the links assigned to its parts; this corresponds to a traversal down one level of the structural hierarchy of that object. Conversely, the links could be assigned to several widely-separated proto-objects, forming a group that would correspond to a (coarsely-coded) object. Thus, even though attentional capacity may be limited, the ability to quickly traverse a structural hierarchy enables attention to rapidly access any part of an object's structure in the image [7].

The challenge, then, is to create "active" displays that output visual information in a manner that matches this style of information pickup. Important factors in this regard are the structural levels that exist in the descriptions of objects and events, and the kinds of information typically present at each level for a given situation.

4.1.2 Graphics

Although graphics is nearing the point where scenes can be rendered with arbitrarily high degrees of fidelity to the real world, such fidelity often comes at the cost of considerable computational effort. Although this cost might be acceptable for creating single images, it is much less so for animations, and may always be prohibitive for interactive graphics, no matter how fast machines will become [20]. As such, it becomes important to determine what parts of an object or event should be rendered in any given approximation (e.g., a given number of polygons). If the goal is to achieve the greatest degree of perceived realism, then human perception must be taken into account.

One approach is to use eye movements to determine which parts of an object are of greatest interest to a viewer, and then provide relatively greater detail to those parts [21]. While this is a good start, this technique has several limitations. First, it is not sensitive to the structural levels that exist—if an eye is gazing at the leg of an animal, it cannot be determined whether the viewer is perceiving the entire animal, just the leg itself, or even perhaps some part of the leg, such as the knee. Second, this technique cannot provide information about the amount and type of information being picked up by the viewer. All it can do is determine spatial location.

Given the model of visual processing outlined in Section 3, the possibility arises of an approach based on attentional allocation rather that eye position. Techniques such as the flicker paradigm (Section 2) may be of use here. An interesting example of flicker-induced change blindness involves an airplane with an engine that appears and disappears [1]. Although all observers immediately see (and therefore attend to) the airplane, most require a considerable amount of time to see the change in the status of the engine. What appears to be happening here is that observers use an initial (or *entry-level*) description of the airplane when they first perceive it, and that a description of the engine is not included at this level. Only later does attention traverse down the structural hierarchy to focus on the engine as an object in its own right; once this occurs, the change is readily seen.

More generally, it is possible to determine which aspects of an object are seen to change most quickly (and are therefore at a level closer to the entry level), and which aspects are seen to change much later (and are therefore at a level further away). Systematic application of such techniques can therefore potentially determine the structural levels used by an individual when they perceive any given object. These structural levels would then specify a set of "natural " levels of detail to select from when rendering an object.

Note that this approach is not restricted to well-defined objects, but could also be applied to regions of the image. This would be of particular importance for "active graphics" (e.g., gaze-contingent rendering [20]), in which most of the detail is allocated to the object(s) being looked at, and minimal detail allocated to the background. There is evidence that sets of items can form large-scale "constellations " that can in some ways be treated as single proto-objects [22]. Moreover, it appears that what is picked up by attention are particular aspects of the distributions of properties within each group, such as average item size [23]. An interesting direction for future work is to explore this issue more thoroughly: for example, determine the sensitivity to changes in averages (or other statistics) of various properties. The results could provide useful information as to the appropriate level of detail for a background seen "at a glance " ; it may well be that relatively coarse descriptions will suffice.

4.1.3 Interface Design

The techniques and theoretical framework described here can also provide guidelines for applications such as visualization and interface design. For example, if change itself is used to help visualize a particular situation, the involvement of attention means that severe limits exist on what the user can perceive regarding the change. In such a case, precautions should be taken to ensure that the display provides the user with appropriate external information to compensate for the limits on the information that can be held internally [24].

[1] A QuickTime movie containing this example can be found at http://www.cs.ubc.ca/~rensink/flicker/download/

More generally, given that attention can access only 4-5 items at a time and only a few properties of those items, care must be used in the design of an interface that will convey information to the user in the most effective way possible. To the extent that the basic graphical items in the display (e.g., text, symbols, etc.) do not map onto distinct "units" of attention, greater time and effort will be required, since:

(i) the proto-objects encoding these graphical items must be broken down into such units.

(ii) some of the units originating from different graphical items are likely to be similar. If so, concurrent access by all attentional links cannot take place if confusion is to be avoided. Instead, attentional access must proceed one item at a time.

Consequently, if information is to be conveyed to the user as quickly and effortlessly as possible, it becomes essential to consider what kinds of attentional units would be involved in the key graphical items.

Several studies of attention have begun to map out the set of attentional units by examining the kinds of items in a display that are immediately noticed by an observer (e.g [8], [25]). While useful, these studies have the potential danger that some of the units discovered in this way *attract* attention but do not enter into the final *description* of the item at the nexus. Variants of the flicker paradigm have been developed that avoid this difficulty [12].

Note that although attentional mechanisms and capacities can be determined for an "average" observer, the amount (and perhaps even type) of attentional resources available will likely vary for each individual, being a function of factors such as age and culture. Moreover, the attentional mechanisms actually used for a particular task will likely be a subset of these, depending on the task itself [13] as well as the other tasks being carried out concurrently [26]. As such, interfaces may need to adapt and adapt—in a very dynamic way—to each user if optimal operation is to be achieved. Change blindness studies (e.g. [13]) can help with this, providing guidelines as to what kind of information is used by what kind of user for what kind of task.

4.2 Visual Transitions

4.2.1 Basics

Change blindness makes invisible any (unattended) transition in the image that could potentially intrude into the awareness of an observer. As discussed in section 2, there are many ways of doing this, including making the transitions during a saccade or blink, during an occlusion by an object passing by, or by simply making the transition sufficiently gradual (e.g. by blending) that it does not draw attention.

The potential "invisibility" of image transitions is a double-edged sword: on one hand noninformative transitions (i.e., disturbances) can be eliminated; on the other, informative transitions might be missed. The usefulness of the transitions therefore determines whether a display will either promote or prevent the conditions that create change blindness.

4.2.2 Graphics

Noninformative transitions can often be found in interactive graphics. One source is the "popping" that can occur due to a change in the level of detail in a rendered object. A similar effect occurs in active graphics, where the contents of a display can be suddenly altered in response to some action of the viewer (e.g., a shift in gaze direction, or a mouse event).

It has been suggested that such transitions be made during saccades so that they will not be noticed [20]. However, any other technique that creates change blindness could also be used for this. Saccades and blinks have the advantage of being constantly-occurring events that are part of normal viewing, and so go largely unnoticed by the vast majority of viewers. As such, they create "invisibility" in the most natural way possible. Although they can be harnessed in a passive fashion by monitoring the viewer and then waiting for a saccade (or blink) to occur, they can also be actively induced: an involuntary saccade can be induced by having something interesting suddenly appear in part of the display; an involuntary blink can be induced by a sudden noise [2]. Note that these active variants not only allow a greater degree of control over when the saccade or blink is made, but also allow these techniques to be applied to multiple viewers.

If it is not possible to make invisible all the transitions in a display, it would at least be useful to maximize the number of transitions that occur simultaneously. Although the motion signals accompanying these transitions will still be noticed, attention will only be drawn to one of the locations involved, thereby minimizing the disruptive effects caused by the other transitions.

4.2.3 Interface Design

Work on change blindness shows that any kind of irrelevant disturbance may potentially interfere with the ability of the visual system to access external information; moreover, users will be largely unaware that such interference is occurring, and so will not attempt to compensate. Consequently, interfaces must attempt to minimize visual events in all aspects of their operation. This is especially important in animated displays in which change itself is serving to convey information: if a disturbance occurs during such an event, valuable information could easily be lost without the user knowing it [23].

With this in mind, several suggestions can be made regarding the design of displays robust to change blindness. To begin with, displays should be such that saccades are minimized. This could be achieved by, e.g., keeping important sources of information relatively closer together. (All other factors being equal, of course.)

In addition, displays should minimize the number of dynamic events occurring in the background, since these will inevitably divert attention from the primary information source. Importantly, displays should also minimize the number of dynamic events in the foreground. Although a single event can be attended without problems, two cannot [3], [12]. Thus, a second dynamic source of information is effectively a distractor, and so will increase the likelihood of change blindness. For optimal operation, then, interfaces should have at most a single dynamic source of information.

Note that given the potential sensitivity of information pickup to visual disturbances, it may be worthwhile to develop methodologies to evaluate the degree to which change blindness might affect a given interface. The techniques developed in various change blindness experiments would be quite useful in this regard.

[2] Interestingly, these techniques have long been used to create various effects in films [27].

4.3 Attentional Coercion

4.3.1 Basics

Given that our visual experience of the world depends on the careful coordination of attention, the possibility arises of the display taking control of attentional allocation and making the observer see (or not see) any given part of the display. Such *attentional coercion* has been used by magicians for centuries to achieve a variety of striking effects [28]. By controlling the systems that allocate attention, it may be possible to put a similar kind of magic into visual displays.

Attentional coercion could be carried out in a number of ways:

(i) high-level interest. Voluntary, high-level control of attention can be "co-opted" by semantic factors, e.g. stories that interest the observer in the particular objects or events.

(ii) mid-level directives. These are cues that cause attention to reflexively move to a given location. Examples are the direction of eye gaze (in a viewed figure), and finger pointing [28].

(iii) low-level salience. Attention is also reflexively drawn to items with "salient" properties—e.g., items with a unique color, orientation, or motion signal [25].

These techniques have proven quite effective when carried out by humans, either as conjurers [28] or filmmakers [27]. Given that machines can have " superhuman" control over the stimuli presented to an observer, the potential exists for effects even more powerful than those currently known.

One possibility is to present items so briefly that they are not consciously registered [16]. Although such "subliminal" presentations may not affect an observer's conscious perception, directives or salient items presented in this way might cause a nonattentional system to direct attention in the desired way. If so, the result would be that the observer would experience nothing out of the ordinary, but would simply "see" that item.

4.3.2 Graphics

High-level coercion has been a mainstay of films for many years [27], and as graphics becomes more sophisticated and more like filmmaking, it will need to make increasing use of elements of storytelling, pacing, etc. to engage the viewer ' s interest and make sure that critical events are not missed due to attentional wandering.

Mid- and low-level mechanisms could also be useful, even though these would likely not be able to result in sustained coercion. Even a transient effect might be useful: If attention could be briefly sent to a particular part of the display at the exact time that a visual transition occurred elsewhere, it might be possible to make that transition invisible (Section 4.2) without any need for monitoring the viewer or inserting any extraneous events in the display.

If such an approach proved successful, this would give rise to several interesting possibilities. For example, if coercion could make transitions invisible without affecting any other aspect of perception, a sudden change in the display would be experienced by the viewer as simply "happening". Since the motion signals that would normally accompany such a change would not be seen, this might be perceived as a "supernatural" event.

4.3.3 Interface Design

Given the importance of attention for the successful pickup of information, the ability to direct a user ' s attention to the appropriate item at the appropriate time would be extremely useful. A coercive display could ensure that important events would not be missed, and might even speed up overall operation directing attention to required locations or items.

Coercion might also provide a useful way to notify a user of a new event that has occurred (such as the arrival of email). Current systems notify the user by a "hard" warning — a noticeable alert that grabs attention. However, notification might be done in a less disruptive way by a "soft" warning, which would simply direct the user ' s attention to the relevant announcement when they were in an appropriate state (e.g., just finished reading an interesting section of text). In such a situation, the user would notice nothing unusual, and the announcement would simply appear, as if by magic.

4.4 Nonattentional Pickup of Information

4.4.1 Basics

For the most part, visual displays have been concerned with the attentional (conscious) aspect of perception. Although preattentive processes at early levels are sometimes taken into account, these are usually considered as providing input to attentional processes, which then construct the representations underlying " real " perception.

However, in the view developed here (Section 3.3), the attentional system is just one of several streams that take their input from early vision: Other, nonattentional systems exist that operate in tandem with the attentional one, and these streams carry out a significant (albeit nonconscious) part of perception. Although these streams are poorly understood at the moment, they are capable of having an effect on several aspects of behavior (Section 3.4). There is consequently potential for new kinds of effects in displays that can harness these processes.

4.4.2 Graphics

The traditional *raison d ' etre* of graphics has been to produce a conscious visual experience in a viewer. But given that nonattentional (and presumably nonconscious) processes are also affected by visual input, the interesting possibility arises of inducing effects that affect a viewer but are not experienced in a direct way (i.e., as a conscious visual image).

Although this possibility is highly speculative, such effects can potentially be achieved in at least a few different ways. One approach is to use active graphics (e.g., a gaze-contingent rendering system) to change background items away from the object that viewer is looking at. If this can be done without drawing attention (Sections 4.2 and 4.3), the only streams that would respond would be nonattentional [3]. Thus, for example, if the gist or layout of an image were continually changing, the viewer might not see this, since attention would not be involved. However, the viewer might still have a "feeling" of something odd happening; this might be experienced as a "sixth sense". Alternatively, mindsight (Section 3.4) may eventually prove to be a reliable way of creating such experiences.

[3] This differs from the background changes discussed in Section 4.1, which involve attentional mechanisms.

4.4.3 Interface Design

Like graphics, interface design has traditionally focused on the attentional aspects of perception. However, just as in the case of graphics, new possibilities may arise from considering the use of nonattentional systems.

For example, the control of actions such as reaching and grasping appears to be mediated by nonattentional systems (Section 3.4). This raises the possibility that activities such as moving a mouse are also mediated by such systems. If so, it may be possible to develop displays that will help a user move a mouse to a given location more quickly. In such a situation, the user may not be aware that the display is providing such guidance; the user simply "does the right thing".

Another interesting possibility involves the "sixth sense" described above, which—if it could be invoked—would likely involve nonattentional mechanisms. If such an effect did not disturb attentional control, this would make an extremely useful form of alert; it would essentially be a second type of soft warning . Such a system could, for example, alert the user to the arrival of email while simultaneously allowing them to monitor a changing display (or engage in any other attention-demanding task) without any degradation in performance.

5. ACKNOWLEDGEMENTS

The author would like to thank Nissan Motor Co. (Japan) and NSERC (Canada) for supporting much of the work upon which this paper is based. Thanks also to an anonymous reviewer for the suggestion that change blindness experiments might be adapted to evaluate the susceptibility of interfaces to change blindness.

6. REFERENCES

[1] Rensink, R.A., O'Regan, J.K., & Clark, J.J. To see or not to see: The need for attention to perceive changes in scenes. *Psychol. Sci*., 8, 368-373. 1997.

[2] Rensink, R.A. Change Detection. *Ann. Review of Psychology*, 53, 245-277. 2002.

[3] Rensink, R.A. Seeing, Sensing, and Scrutinizing. *Vision Research*, 40, 1469-1487. 2000.

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

[5] Kahneman, D., Treisman, A., & Gibbs, B. The reviewing of object files: Object-specific integration of information. *Cog, Psych.*, 24, 175-219. 1992.

[6] Wolfe, J.M. Guided search 2.0. *Psychonom. Bull. & Rev.*, 1, 202-238. 1994.

[7] Rensink, R.A. The Dynamic Representation of Scenes. *Visual Cognition*, 7, 17-42. 2000.

[8] Rensink R.A., & Enns J.T. Early Completion of Occluded Objects. *Vis. Research*, 38:2489-2505. 1998.

[9] Wolfe, J.M. Inattentional amnesia. In V. Coltheart (Ed.), *Fleeting Memories*. (pp. 71-94). Cambridge, MA: MIT Press. 1999.

[10] Pylyshyn, Z.W., & Storm, R.W. Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision* , 3, 179-197. 1988.

[11] Levin, D.T., & Simons, D.J. Failure to detect changes to attended objects in motion pictures. *Psychonom. Bull. & Rev.*, 4, 501-506. 1997.

[12] Rensink, R.A. 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. 2001.

[13] Ballard, D.H., Hayhoe, M.M., Pook, P.K., & Rao, R.P. Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20, 723-767. 1997.

[14] Merikle, P.M., & Joordens, S. Parallels between perception without attention and perception without awareness. *Consc. and Cog.,* 6, 219-236. 1997.

[15] Goodale, M.A. Visual routes to knowledge and action. *Biomedical Research*, 14, 113-123. 1993.

[16] Marcel, A.J. Conscious and unconscious perception: Experiments on Visual Masking and Word Recognition. *Cog. Psych*., 15, 197-237. 1983.

[17] Clancey, W.J. *Situated Cognition* (pp. 101-132). Cambridge: Cambridge University Press. (1997).

[18] Clark, A. *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press. 1997.

[19] Marr, D. *Vision*. San Francisco: Freeman. 1982.

[20] O ' Sullivan, C., Dingliana, J., & Howlett, S. (2002). Eye Movements and Interactive Graphics. In *The Mind's Eyes*. Hyönä, J. Radach, R. and Deubel, H. (Eds.) Oxford: Elsevier Science. (To appear, 2002).

[21] Janott, M., & O ' Sullivan, C. Using Interactive Perception to Optimise the Visual Quality of Approximated Objects in Computer Graphics. In *10th Eur. Conf. on Eye Movements (Abstr),* 110-111. 1999.

[22] Rensink, R.A. Differential Grouping of Features. *Perception*, 29(suppl.), 99-100. 2000.

[23] Ariely, D. Seeing sets: Representation by statistical properties. *Psych. Sci*., 12, 157- 162. 2001.

[24] Nowell L.T., Hetzler, E.G., & Tanasse, T. Change Blindness in Information Visualization: A Case Study. *Proc. IEEE Symposium on Information Visualization* 2001 (INFOVIS '01), 15-22. 2001.

[25] Treisman, A. Preattentive processing in vision. *CVGIP*, 31, 156-177. 1985.

[26] Pashler, H.E. *The Psychology of Attention* (pp. 265-317). Cambridge, MA: MIT Press. 1998.

[27] Dmytryk, E. *On Film Editing*. Boston, MA: Focal Press. 1984.

[28] Sharpe, S.H. *Conjurers' Psychological Secrets*. Calgary, AB: Hades Publications. 1988.