In C. Taddei-Ferretti (Ed.),
Neuronal basis and psychological
aspects of consciousness, in press.
Singapore: World Scientific.

Very rough, many imprecisions. Please don't quote; comments, on the other hand, are welcome.

ATTENTION: EXPORTING VISION TO THE MIND

PATRICK CAVANAGH
Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, USA

ABSTRACT

Attention is widely seen as a limited capacity process which selects relevant streams of input and integrates their piecemeal features into coherent object descriptions. The resulting internal model of the scene is the content of our visual awareness and forms the basis of the stability of our experience of the world. I describe a different view where attention is not the final stage but just one direction taken by visual information, a direction which leads specifically to a sparse conscious model of the world. The role of attention is seen as translating visual descriptions generated by the primary visual system into the highly constrained public code of conscious perception. The properties of the public visual code as assessed through conscious report of visual phenomena are therefore not necessarily good indicators of the complexity of descriptions available in the primary visual system.

1. Introduction

The study of attention has a very long history. Hatfield (1996) points out that Aristotle described the single focus of attention (3c BCE), Lucretius the aspects of active selection and improved sensitivity (1c), and Augustine of Hippo added a description of involuntary shifts of attention (4c). More recently, Alhazen (11th century) noted that scrutiny is required to see fine detail whereas other features are visible (pop-out) at first glance (Sabra, 1989). Wolff proposed a theory of feature integration in the 18th century, suggesting that things are combined into wholes through attention (Hatfield, 1996).

There has been a great wealth of new studies of attention in this last century but no new additions to this list of the properties of attention. In general, the selective filter of attention is often taken to be the last step in perception, feeding integrated representations into the final model of the world which fills our visual awareness. As the gateway to awareness, it is also the spillway to memory, at least explicit memory. There has been constant argument about the level to which information is processed when it is not selected by attention but the general consensus is that attention is the final step or bottleneck in visual processing (Figure 1). The fate of unattended information can be examined indirectly with among others, priming tasks, adaptation effects, or automated motor responses (grasp, walking). These effects are often seen as curiosities — the intriguing end states of the chaff of mental processing. It may be highly processed chaff in some cases, but chaff nonetheless.

I would like to suggest a different view, one that reorganizes the elements of vision, attention, and conscious perception in a different structure. I will propose that attention is not the final stage in the integration of a coherent visual experience but just one direction taken by visual information, a direction which leads specifically to conscious perception, a condensed product for export to the rest of the mind. There are other equally important paths taken by visual information and the content of conscious perception should not be taken as a true indication of the limits and properties of visual representation. I will call the core visual system which services all of these clients the *primary visual system.* These ideas are closely related to those of many others, most clearly to Baars (1988) concerning the public broadcast metaphor of consciousness and to Goodale and Milner (1992) concerning the role of alternative visual systems whose content is unavailable to awareness.

This paper begins with a review of the central properties of attention as currently understood. It then sketches out very roughly the nature of the organization that I am proposing.

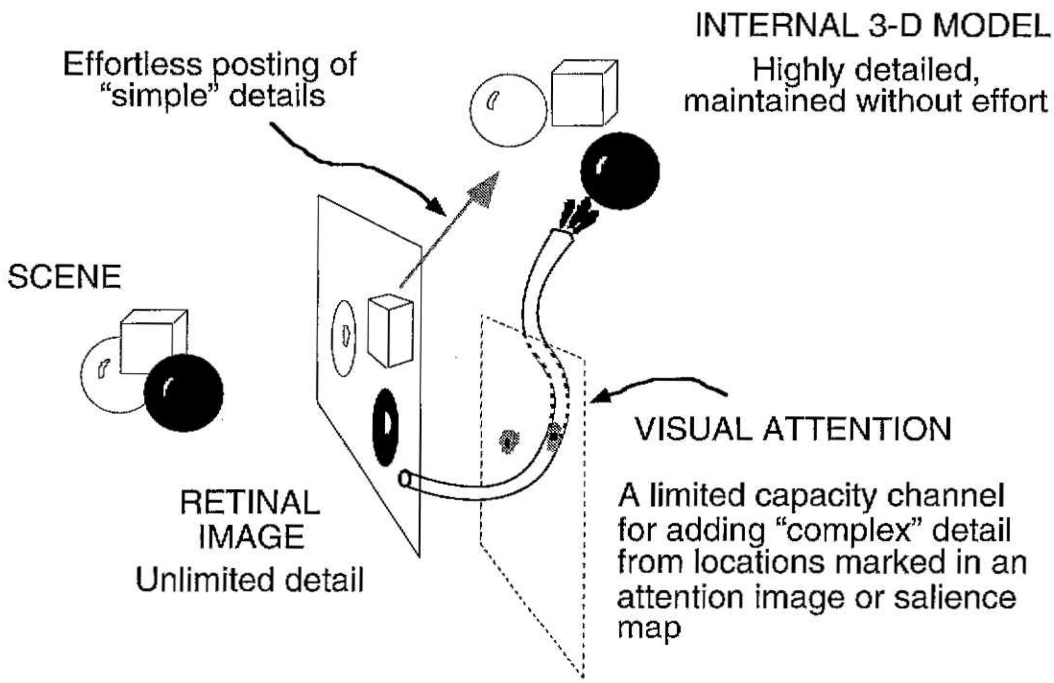

Figure 1. A widely accepted view of vision supposes that a stable representation of the 3-dimensional world is built up by adding simple details effortlessly and integrating into those, more complex details with the help of attention. Once constructed, this internal model of the world is considered to persist without additional effort. This paper describes several challenges this view and suggests an alternative.

2. Attention and selection

Many have noted that attention to a visual event provides greater clarity in discerning the details of the event. This extra resolution afforded to attended items has been measured in a number of ways. Figure 2 shows just one example. It is easy to notice that 3 of the items are lighter than the others while looking at the central plus sign. On the other hand, unless we attend to each item in turn (again without moving the eyes from the plus sign), we cannot distinguish which ones are slightly wider than the others. This conscious scrutiny supplies the additional resolution that allows us to discriminate the slight deviation from a square shape in each. While we are consciously analyzing each target, the exact shapes of the others are not readily available to us.

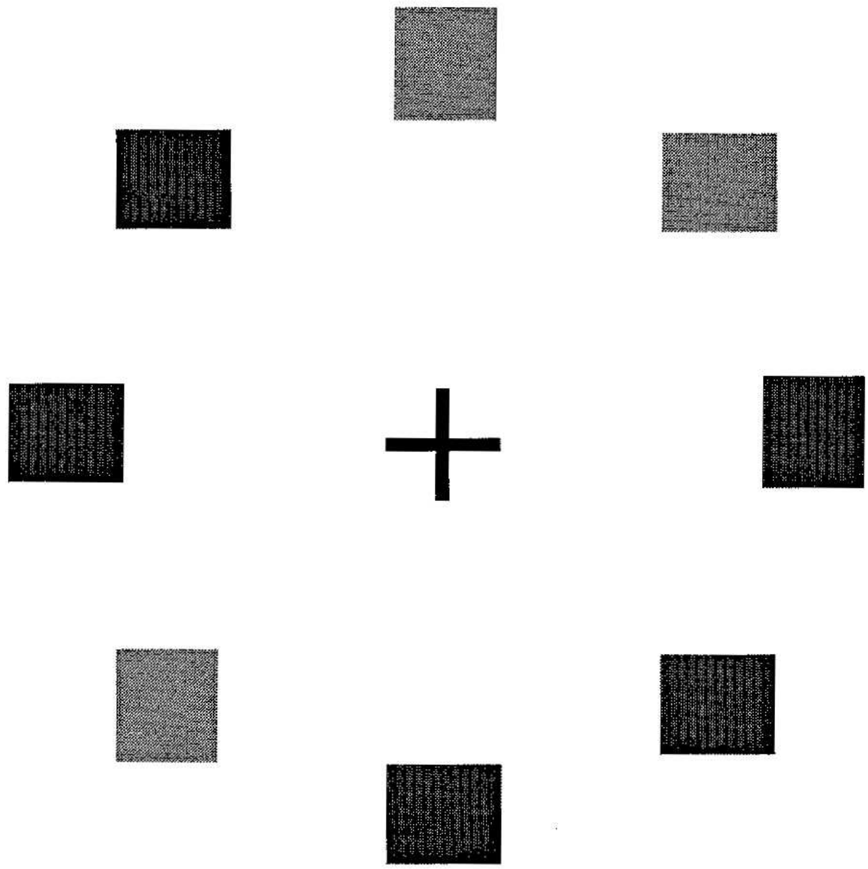

Figure 2. Attention is not required to notice the simple feature of lightness — the 3 lighter rectangles stand out from the darker ones. Identifying the rectangles which are higher than wide is more difficult. The difference is to small to resolve without further effort. Keeping the eyes fixed at the center, attention must be directed to each rectangle in turn, one at a time, in order to make the distinction.

This aspect of selection has been the defining feature of attention since the earliest writings on the topic. Attention has been seen as selecting a source of interest among competing sources. Although the purpose of this selection is to focus on the relevant material, it is also taken as a consequence of a limited processing capacity — only so much can be accomplished so the available resources are allocated where they are most needed. No particular explanation is offered for why capacity might be limited other than the circular comment that we could not really process everything, there is just too much.

If most have agreed that selection is a key role for attention, there has been less agreement on the fate of material which is not selected. This controversy has been phrased as an early versus late selection debate. In the case of early selection theories, material not selected by attention is assumed to undergo no further analysis. In the late selection theories, several streams of input are assumed to be analyzed to a high level and selection focusses on one highly processed signal while the others are just ignored.

There are a number of phenomena which support some degree of late selection. The most obvious example comes from the common experience of attending or selecting a conversation of interest to monitor in a crowded noisy room. The ignored conversations seem to drop out of our awareness — except if our own name comes up and we seem to instantly tune in. Clearly, we could not respond to our own name if we had not been processing the ignored conversations to some extent.

An even more dramatic example comes from the recent work of Mack and Rock (1996) where they report on a phenomenon they have labeled Inattentional Blindness. In their studies, subjects were asked to perform a simple detection task at a known location slightly off the center of gaze. They proceeded as expected through, say, the first 2 trials of reporting whether a dot was present at that location or whether a shape presented there was wider than it was high. On the third trial, an unexpected word was presented right at the center of gaze. The experiment was stopped and the subjects were asked if they noticed anything different on the last trial. About 95% of the time, subjects claimed not to have noticed *anything*. Despite this blindness to the unexpected, unattended event, a subsequent priming task showed a strong effect of having been presented with the word. A stem completion task was used where, say, the first 2 letters of a 5 letter word (e.g. CO_ _ _) were shown. Typical completions of the stem were collect from several subjects in a separate group (couch, count, etc.) and these rates were compared to those for the subjects who had been in the 3 trial detection condition. These subjects had many more completions using the word they claimed not to have noticed. Clearly, even an event which goes completely unnoticed may be processed to a very high level, as high as at least the shape of a word.

A final example can be taken from physiological studies of binocular rivalry (Leopold, 1997; Logothetis & Schall, 1989). In these studies, different images a presented to the two eyes at the same time. Typically, if the images are overlapping and small enough, only one of them is perceived at any moment. In addition, there is an alternation between the two as they switch back and forth into awareness. This binocular rivarly can be studied at the level of individual cells in the visual system and in some areas of the cortex, some cells are changing their activity in synchrony with the behavioral report of the animal (and so with the assumed conscious percept of the animal). However, there are also other cells which show the opposite behavior, responding as if they were coding the unseen pattern. Although much work remains to be sure of the exact role of these cells, it is plausible that they represent the parallel analysis of the alternative, unselected percept.

This array of evidence for complex analysis of unselected information raises the question of whether "selection" for the final levels of visual analysis is the appropriate metaphor for the processing which is occurring for attended events. It appears that some events are processed at least as deeply without attention as they might be with attention. Indeed, the difference appears to be mainly whether the observer can report the content of the event — whether there is an episodic memory (chronological event memory) laid down for the event. I will argue that the selection by attention is not the final stage of vision, but is a separate branch of output for visual information, one that is intended for "public broadcast" and one that has its own constraints which are not necessarily representative of the overall visual analysis of events. Since most experiments on vision are based on a subject's reports (i. e. memory) of their conscious percepts, the large body of information concerning vision has mainly tapped only this separate branch of conscious perception, the experience of attended and therefore remembered events.

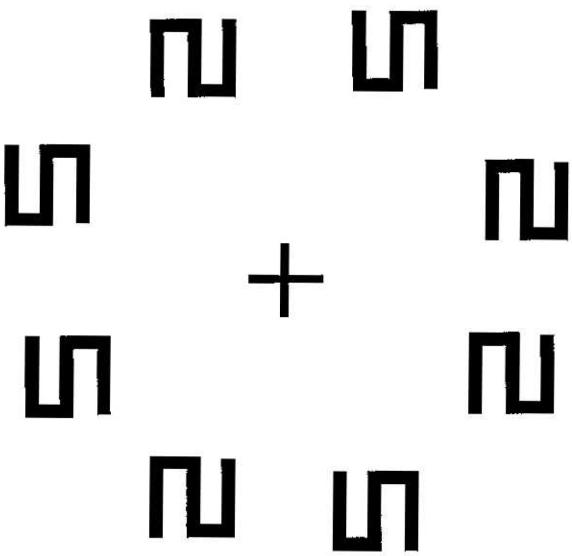

Figure 3. Attention also integrates the elements of each pattern. The difficulty in locating the u-n patterns among the n-u patterns is not one of resolution. Turn the display 90 degrees and the distinction becomes trivial, showing that all of the necessary detail has been resolved. What is difficult when the shapes are not identified as familiar patterns is coding the relative locations of the line segments — integrating the parts into the whole (adapted from Wang, Cavanagh, & Green, 1994).

3. Attention and feature integration

Increased resolution is not the only advantage accorded by attention as demonstrated in Figure 3. The different patterns are hard to distinguish when fixating the central plus sign; each has to be scrutinized in turn to see which one is a u-n and which an n-u. But this ability to distinguish the items with the help of attention is not because of extra resolution that is afforded, but because attention is required to piece together the relative positions of the line segments. The evidence that extra resolution is unnecessary is found by rotating the page 90 degrees. The two types of pattern are now clearly distinguished without individual scrutiny because they now correspond to familiar digits and there is no need to specify the relative positions of the lines to differentiate one from the other (Wang, Cavanagh, & Green, 1994).

According to some, this integration or binding of the separate features of an object into a coherent whole is a central function of attention. This was first proposed by Wolff in the 1800s (Hatfield, 1996). More recently, Treisman and her colleagues (Treisman & Gelade, 1980; Treisman & Gormican, 1988) have added a rich assortment of experimental evidence to this proposal. In her view, the visual array is broken down into a large set of features, and the spatial pattern of the instances of each feature type is given on a separate map, say one for red, one for green, one for near, one for far, one for vertical, etc. Each feature which has its own map is termed a primitive feature and the presence of such a feature in the visual field can be ascertained for the whole visual field at once — a so-called pop-out of a primitive feature. A single red dot will stand out from a background of green dots, for example. On the other hand, if an object has several primitive features, attention must be directed to the object's location in order to bind the features together. To know that an item is both red and horizontal, for example, requires that attention be directed to each item in turn, checking to see if both features are present at that location. This conjunction search is consequently much slower than the feature search.

Several studies have challenged the claim that conjunctions require attention to be identified. In one such example, red vertical bars can be quickly detected in a field of red horizontal and green vertical and green horizontal bars. The target cannot be found by searching for any one feature, say red or vertical, because these features are also present in the display in other combinations. One possible explanation proposes an alternative mode of searching for targets wherein the subject pays attention only to the group of items having one feature — say all the red ones. Within the group of red bars, the desired target is the only vertical bar and so it stands out as if the green items were never there.

The possibility that attention is required to bind features together therefore still stands. Certainly there is no other explanation for the difficulty of distinguishing items in Figure 3 above nor those in one of the more difficult conjunctions — adjacent pairings of colored rectangles where the odd item has the reversed spatial order (Figure 4, Wolfe et al, 1990). Here it is again very difficult to locate the item which has, say, the darker rectangle on the left as opposed to the right. There is now no way to group the items by any individual feature since every items has both a light and a dark vertical bar. Each item must be scrutinized in turn.

Nevertheless, it is reasonable to ask how we are able to move around in the world without being overwhelmed by the processing requirements of all the conjunctions we must respond to. Red traffic lights are red, round, and up above ground, chairs are flat and horizontal but lower than desk tops. And yet we stop almost automatically and sit without a second thought on chairs only briefly glimpsed. I will show in a later section that the difficulty in dealing with conjunctions may be a problem only of conscious perception, not one of primary vision.

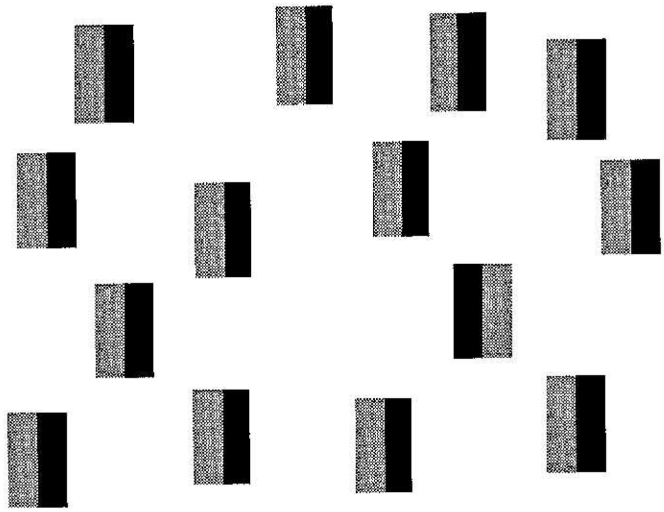

Figure 4. The target is the only pattern with its dark bar on the left but this is very difficult to locate. Each pattern has to be scrutinized in turn. The binding of the features into a single object appears to require attention.

4. Conscious perception is sparse

It might not seem like much of a limitation that our understanding of vision deals mostly with conscious perception. After all we see a rich and detailed world. What else is there? A few streams of unattended but highly processed information do not seem to be of much consequence if they are simply more of what we are seeing but not at the moment available to awareness. However, recent work has dramatically altered how we think of the consciously experienced world. Rather than rich and complete, it seems that it is sparsely represented and constantly shifting.

Many agree that attention is the gateway to conscious perception. Recent evidence for this is increasingly strong. As mentioned above, Mack and Rock have demonstrated that events which would otherwise be easily seen go unnoticed if they are unexpected even when presented at the center of gaze. Joseph, Chun, and Nakayama (1996) have shown that events which have been claimed to require no attentional processing — the 'pop-out' of a tilted line among vertical lines — go undetected if attention is diverted.

But what of our feeling that we piece together our world in multiple glances, building up a reasonably complete model of a stable world around us? This concept of a rich model of the world does not hold up. Rensink and his colleagues (Rensink, O'Regan, & Clark, 1996) have shown that large dramatic changes to images go unnoticed unless they occur to objects currently under scrutiny by attention. A picture of a stream with a windmill and a bridge across the stream is presented followed by a quarter-second blank interval and then the same scene again but now the bridge reaches only half way across the stream. This alternation of images is repeated until the subject notices the difference between them. It can take up to a minute for people to notice this very large change (the blank interval suppresses the motion transient which would normally call attention to the change). They notice it right away if they were paying attention to the bridge when the first image was presented. Similarly, McConkie and Curry (1996) have allowed people to scrutinize an image of say, a house with a driveway. Every time the subject makes an eye movement, the authors would change something in the scene. Most of the changes go unnoticed. Simons and Levin (1997) have made the same point with scene changes in video and movie editing. As the camera angle changes from one view to another, a change in clothing or an object or even replacing one person with another goes unnoticed by the majority of observers.

These changes would not be missed if we did have a model of our surroundings that was as complete as we feel it to be. We would be very disturbed if, in fact, the world changed every time we blinked, if someone new was substituted for our waiter every time he or she appeared. But these studies have demonstrated that we would not notice these changes even if they were occurring regularly. Luckily the world is not so instable but this stability is a property of the world, not our representation of it.

Not only is attention necessary to select the elements for this sparse conscious description of the scene but it appears that it is required to maintain the descriptions as well. Wolfe (1997) used several tasks where attention would analyze particular elements in an array and then move on to others. He found no evidence of persisting benefits. Any further calls on the identity of the once scrutinized object required new scrutiny of the same magnitude as the first time. Wolfe described post-attentive vision as having the same disorganized state as pre-attentive vision. His results suggest that it is not enough for attention to have analyzed an object for it to be added to an accumulating 3-D description of the world. Attention is required to keep that description available. If attention wanders off elsewhere, the items so laboriously constructed are no longer available to conscious inspection.

O'Regan has made the strong claim (1992) that there is only the most minimal internal representation of the visual world, that the world itself serves as its own representation — after all, it is always there and if we want to know something about it, we just go look. The idea that the world is the external memory for vision is based on the point that, unlike, say speech sounds, the world does not go away. The same could be said of pain. We do not need an internal memory for where our body hurts. Say that some weasels were gripping onto our body at a few different spots. We would not have to ask ourselves "now where are those weasels?" However, I would apply his description of minimal representation only to the consciously perceived world. Our conscious abilities are flexible and powerful and can scrutinize each area of interest as needed to fill in information. This vision-on-demand gives the impression that everything is maintained as a conscious representation when in fact very little is maintained and none of that is maintained without continued effort.

If all this rapid, mental handiwork is required just to create a *false* sense of "knowing what is out there", we can again ask how is it we so effortlessly and accurately navigate around our world, sitting down on chairs, picking up forks, driving while talking. Some authors have suggested that the motor systems have their own simple visual representation of the world and that this is sufficient to support simple motor behavior (Goodale & Milner, 1992). I will argue that it is conscious perception which is the supplementary, condensed visual system. The degree of representation within the primary visual system itself (as opposed to conscious perception) may be much more extensive, affording the support for motor behavior and the descriptions of objects and events which can be picked up rapidly by conscious perception when the need arises.

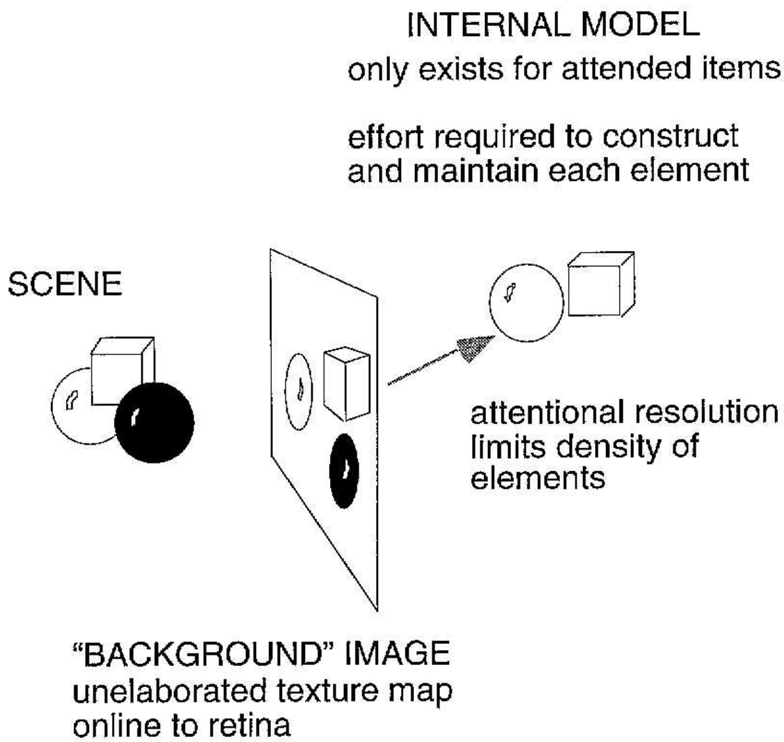

Figure 5. Evidence that many changes in the world around us can go unnoticed suggests that the internal model of our visual world is very sparse. Changes in background, unattended elements in a scene are difficult to detect. Moreover, elements only remain in the sparse internal description if attention continues to be allocated to them. If attention moves elsewhere, the integrated description constructed by attention is no longer available.

As an interesting side issue of this proposal for a sparse conscious experience of the world, we can ask what we experience in the gaps between currently attended objects? It is possible that between the richly described attended objects is simply a detailed 2-D texture corresponding to the retinal image. Remember that the 2-D array that falls on the retina is hardly changed at all by subsequent analyses — brightness and colors are normalized, but the 2-D position of any particular image feature is rarely affected. However, I will argue in below that the detail filling the gaps is a very compressed description of the primary analysis of the scene.

5. Attention and conscious perception

It is not clear what the boundaries are between attention and consciousness, if any. Baars (1996) argues that they are distinct, that attention is selection and consciousness is experience. He makes the analogy to watching television where the act of selecting a channel is quite separate from the conscious experience of the television show. However, this distinction breaks down in light of the recent work by Wolfe (1997) showing, I argue, that attention is not only necessary to select the content of awareness but also to maintain its availability. Once attention is diverted, the descriptions which have been constructed are no longer available. If both selection and maintenance are within the role of attention, it is hard to identify what other processes need to be added to specify the central processes of conscious perception. Certainly it cannot be the content of consciousness itself anymore than a river depends on the specific water currently flowing through it. As long as it is water, it is a river. On the other hand, we can always point to the subjective experience of consciousness and claim that it is not a component of attention. I will have to set that question aside for the moment. With that caveat, I have taken the approach in the rest of this paper that the role of attention is to select and maintain the content of conscious perception and that the two are so tightly linked that to study one is to study the other.

6. Exporting vision to the brain

6.1 The public billboard

It is possible that attention has been incorrectly positioned in current views of vision. I suggest that rather than being the final stage in integrating visual descriptions of the world, its role is to support a separate pathway with very different goals, the pathway which mediates conscious perception. Why would this be considered as a pathway separate from primary vision? Conscious perception should be viewed as a description of the visual world to the rest of the brain — it is not necessarily a good representation of the information generated by the visual system, it is only one strongly constrained description of it.

If we accept the relatively modular nature of the brain, wherein the frontal lobes, the motor areas and the sensory regions have a good deal of independence, it is clear that coordinated activity requires some form of communication among these modules. The modules must talk to each other and it is reasonable to assume that each module will make announcements of its events available to all other modules rather than sending specially tailored notes to each (e. g. Fodor, 1983). This is the popular idea (Baars, 1988; 1996) of a public billboard where modules post events, take notice of events posted by others, and make requests for information. Much has been written about this idea, including suggestions that, for example, the content of the billboard is the content of awareness, but I would like to concentrate on two aspects in particular. First the billboard requires a common language for these exchanges, a common format interpretable by all the participating modules. This has been referred to as "mentalese" by some. Second, the bandwidth of the billboard must be limited (a point discussed in more detail in section 6.4 below).

Taking the case of the visual system as an example, I am suggesting that a major function of attention is the translation of representations generated in the visual system into this public language. Some events in the visual representation are structured so as to be very easily and rapidly translated into the public format and posted on the billboard of the mind. Other events require substantial processing to be translated.

In this view, data on visual search, for example, tell us not about how rapidly a visual representation of a stimulus is created but rather, how rapidly it can be translated into the public format. Difficult to locate conjunctions might be rapidly available for other types of responses (priming, control of walking, etc.) which do not require public posting (visual search requires public posting in order to link a particular stimulus arbitrarily to a particular response).

This is not really much different from the late selection approach described in the earlier sections except that the assumption here is that the visual representation in primary vision is highly elaborated without the intervention of attention. Attention serves only to read out requested portions of the representation and a translate them into public format.

6.1.X Scene properties (this section not in original chapter)

A number of studies have already demonstrated that difficult conjunctions can become easy to detect if they are arranged to support "scene-based properties". Most notably, Y shapes of two different orientations are hard to distinguish but when surrounded by hexagons they form cubes of different 3-D orientations and the oddly oriented cube is very easy to pick out (Enns & Rensink, 1990). This might be taken as evidence that some conjunctions of elements are bound together without intervention by attention, as I would argue. However, this result has not been taken to challenge the notion that attention is necessary for conjunctions but rather to enlarge the list of properties which are rapidly computed across the visual field to include some scene properties as well as primitive features. According to this view, some aspects of 3-D structure or scene lighting might be derived at a very early stage either because of their importance or simplicity. This argument cannot stand, however, for it assumes that the visual system can know which pattern elements form "scene properties" before beginning to analyze them. It could only do so if it had also integrated or conjoined the elements of the Y junction, for example, which itself led to slow search. The only simple explanation is again that the elements of the Y junction *are* explicitly conjoined by the primary visual system into an integrated shape but no public code exists to clearly distinguishes this shape in the two different orientations. A slow translation in to some more awkward set of labels is the result. With the hexagons outlining the Y junctions, the shapes now have simple public labels available of cubes in particular orientations. Although this post hoc analysis is suggestive, a more compelling argument could be made if "scene properties" are avoided and an example is drawn from classic 2-D conjunction patterns.

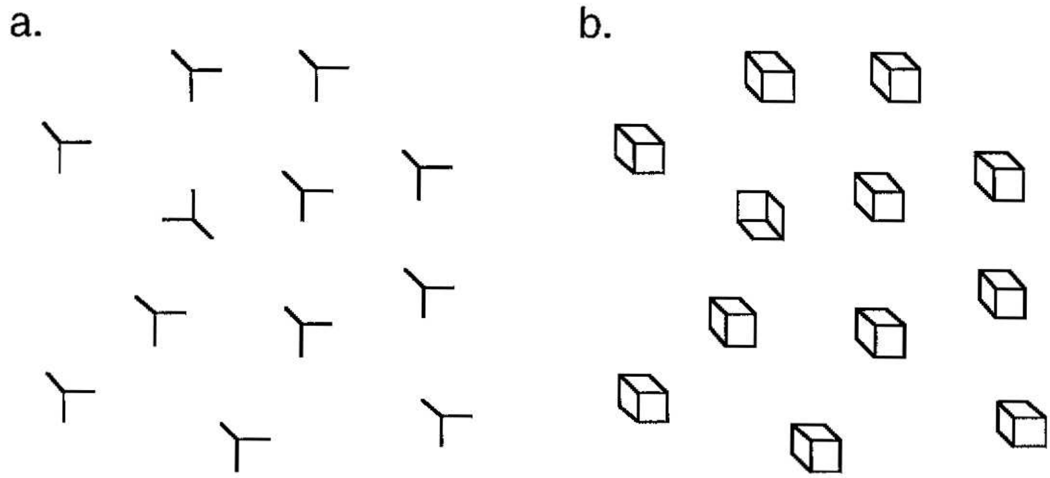

Figure 5x. The Y junctions of different orientations are hard to distinguish on the left but by placing an identical hexagon shape around each, they become cubes and the oddly oriented one is much easier to pick out (Enns & Rensink, 1990). The authors argue that the line elements on the left still require attention to be conjoined whereas on the right, the line elements form 3-D "scene-based properties" and so engage rapid processing. This seems implausible, however, because it requires prior knowledge of which image elements will be "scene properties", a category which itself is ill defined. It is more likely that the elements in both sets of patterns are conjoined in similar manner but that there is a good match between the cubes and simple description in a public format whereas the shapes on the left afford no easy translation into short public descriptions.

6.2 An experiment

An important test of this idea would be to show that conjunctions which are particularly hard to find (Figure 4, for example) are in fact explicitly coded in primary vision. In other words, the conjunction may be hard to find because it is difficult to code the information into the public format even though it is completely described at some level in the visual system. There are many anecdotal examples of this which I have mentioned throughout: being able to drive, walk, or sit down with little conscious direction would be hard in a world only described by primitive feature maps. Nevertheless an example within the visual paradigm would be helpful.

An example is given in Figure 6 below where the elements of the difficult search of Figure 4 are arranged in a different way. This example suggests that the conjunction information is indeed explicitly represented but cannot be easily read out in terms of left/right relative position. In the new organization, the elements are arranged around a contour with the lighter side always facing inward. A single element is reversed and this is relatively easy to find, much easier than the odd item in the array of left/right pairs of Figure 4. The arrangement of the elements into a ragged contour appears to engage some visual process that is sensitive to the light/dark ordering in each pair. This might be a process which segments the scene into regions of different lightness by examining the contrast polarity of disparate bits of border. The one element with the opposite polarity then stands out.

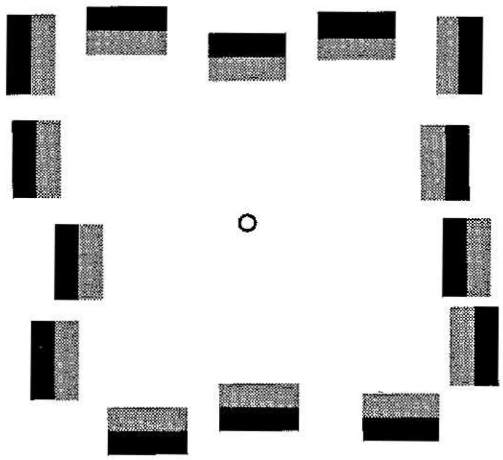

Figure 6. The one patch whose dark bar faces inward is easier to locate here than the one pattern with the dark bar on the left in Figure 4. One possibility is that the primary visual system has discovered an inner, lighter region based on the polarity of the bits of contour surrounding it. The one odd element on the contour stands out. In order for this region-growing process to work, the relative positions of the light and dark bars in each pattern must have been available explicitly. The difficulty in locating the target when the patterns are in a left/right arrangement in Figure 4 could be due to the difficulty in translating the available information about the relative left/right position into a public code. It is not necessarily evidence that attention is required to derive the relative locations of the adjacent light and dark bars.

This ease of performance in this task requires that the relative location of the light and dark pieces be available explicitly without the scrutiny of attention. This suggests that the conjunction of lightness and position that defines the target is coded explicitly at an early stage. Perhaps because no task of ecological importance (prior to reading) required left/right judgments, the public code is ill suited to for the left/right task, but has no problem with an inside/outside task. In other words, the difficulty of locating conjunctions in visual search does not mean that attention is required to integrate features. An integrated description of many visual properties may be generated in the primary visual system without the intervention of attention. It is just that many of these descriptions may not easily be read out in the available "terms" of the public code and therefore have to be laboriously translated into the best approximation under attentional control.

Before we can accept this result, it is important to show that the difficult conjunctions of Figure 4 are not made easy simply by providing some convenient feature-based grouping in Figure 6, as was described for several "easy" conjunctions in section 2 above. The random jitter in locations was introduced for this reason. In particular, isolating the light bars or dark bars alone does not give a suffieciently reliable enough to respond much better than chance and yet responses were 95% correct in the experiment. These results are only preliminary but if they hold up and if further tests of a similar nature continue to support this result then it appears that conscious perception may not always be the most completely elaborated description of the visual world available in the brain.

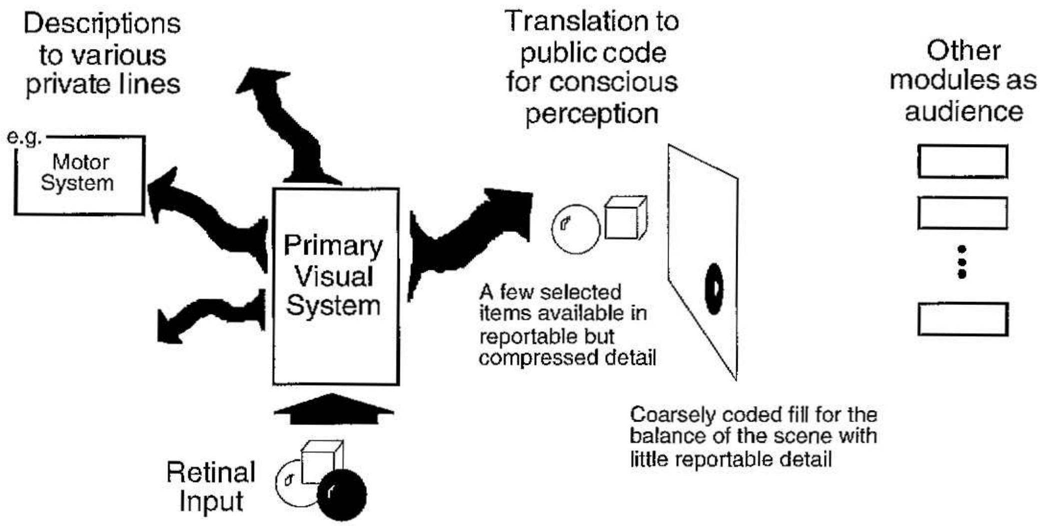

Figure 7. The primary visual system generates elaborated descriptions of the visual world and exports these to various "clients". The description exported to conscious perception must be coded in a public format to be understandable to all the modules connected to the public broadcast. The role of attention is to select descriptions, translate them and maintain them on the public billboard. Many of the properties of vision may be properties of this common code; the representations of the primary visual system may be less constrained. This sketch has only noted the flow of information from the retina to other centers. Information is undoubtedly flowing in the other direction as well but this paper has not addressed the role or format of this flow.

6.3 Private lines

Of course, in addition to public posting, we can assume that private lines can be set up between modules to handle heavy repetitive traffic between two or more modules. Private lines could be thought of as individualized translators between specific modules and these, to carry the metaphor through, apparently run without attention in the sense that they govern automated responses.

6.4 Why is capacity limited?

We can think of the individual modules, like the primary visual system, as having high bandwidth enabling them to represent sensory events richly with very little limit to their capacity. The public billboard, on the other hand, might have limited capacity. Why? The answer is a computational one. If the billboard has to link events in two different modules, for example, it requires a representational complexity equal to the sum of the complexities of each module — it has to be able to post all possible pairs of messages from the two modules. If it were to post messages with all the richness that each module can code, then the billboard might have to be as big as all the modules combined. In some neural models, it would have to be as big enough to code the square of the number of states of each module in order to be able to connect them in all possible ways. The brain would have more space dedicated to the billboard than to the modules themselves.

There are therefore advantages to having a small capacity billboard and (equivalently) a limited language which is shared by all. Physiologically, consider the hippocampus and its associated cortical neighbors. It is much smaller than the cortices whose representations it must combine in order to form memories of conscious events (explicit memory). The language of those memories must be a much reduced version of the neural activity generated by the events in each cortex but it is probably a complete representation of our conscious awareness or at least reportable awareness of the events.

What is the implication of this limited capacity? It is not only that we can accomplish only a few attention-demanding tasks at a time (not enough space to post all these messages) but also that the codes describing the, say, visual events might be much compressed from the richness available in the primary visual system. The conscious experiences of visual events might be coded in compressed or discretized steps along various dimensions as suggested by various compression algorithms used in communications. Some experiments by Wolfe et al (1992) do show categorization effects on visual search for orientation that are along these lines.

7. Attention-based visual routines

It is possible that some levels of description worked out by attention to be posted on the billboard are not available in any form to the primary visual system. In this case, no private line connection could ever emerge for these aspects of visual representation, attention and awareness would always be required.

Simple examples are counting of elements in a display or deciding whether a dot is inside or outside a closed contour (Ullman, 1984). The processes we have studied in most detail in our lab include attentive tracking (Pylyshyn & Storm, 1988) of random and constrained motion displays (like point-light walkers). We have proposed that attention tracking supports an entire motion analysis on its own (Cavanagh, 1992) and studied the parietal activation that accompanies tracking using fMRI techniques (Culham et al, 1997).

8. Conclusions

This paper proposed that one of the foremost roles of attention is to translate the description available in the primary visual system into the formats used for public broadcast to the rest of the mind. Much of the measurement of the deployment of attention may reflect properties of this translation step rather than properties of the underlying visual representation.

Acknowledgement

This research was supported by an NIH grant to PC.

References

Baars, B.J. (1988) *A cognitive theory of consciousness*, Cambridge, UK: Cambridge University Press.

Baars, B.J. (1997) *In the theater of consciousness,* New York Oxford: Oxford University Press.

Cavanagh, P. (1992) "Attention-based motion perception", *Science* 257:1563-1565.

Culham, J. C., Cavanagh, P., Kanwisher, N. G., Brandt, S., Dale, A. M., and Tootell, R. B. H. (1997) "Attentive tracking of moving targets produces parietal activation revealed by functional magnetic resonance imaging", *Investigative Ophthalmology & Visual Science* 38:S1174.

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

Fodor, J.A. (1983) *The modularity of the mind: An essay on faculty psychology*, Cambridge, MA: MIT Press.

Goodale, M.A. and Milner, A.D. (1992) "Separate visual pathways for perception and action", *Trends in Neurosciences* 15:20-25.

Hatfield, G. (1996) "Attention in early scientific psychology", in: *Visual Attention*, R. D. Wright, ed., New York: Oxford University Press.

Joseph, J.S., Chun, M.M., and Nakayama, K. (1996) "Attention requirements in a 'preattentive' feature search task", *Nature* 387:805-807.

Leopold, D.A. (1997) Brain mechanisms of visual awareness, Unpublished Doctoral Thesis, Baylor College of Medicine, Houston, TX.

Logothetis , N.K., and Schall, J.D. (1989) "Neuronal correlates of subjective visual perception", *Science* 245:761-763.

Mack, A., and Rock, I. (1996) "Inattentional blindness: Perception without attention", in: *Visual Attention*, R. D. Wright, ed., New York: Oxford University Press.

McConkie, G. W., and Curry, C. (1996) "Visual stability across saccades while viewing complex pictures", *Journal of Experimental Psychology: Human Perception and Performance* 22:563-581.

O'Regan, J.K. (1992) "Solving the 'Real' mysteries of visual perception: The world as an outside memory", *Canadian Journal of Psychology* 46:461-488

Pylyshyn, Z.W., and Storm, R.W. (1988) "Tracking multiple independent targets: Evidence for a parallel tracking mechanism", *Spatial Vision* 3(3): 179-197.

Rensink, R., O'Regan, J.K., and Clark, J.J. (1996) "To see or not to see: The need for attention to perceive changes in scenes", *Investigative Ophthalmology & Visual Science* 37:S213.

Sabra, A.I., (1989) *The optics of Ibn Al-Haytham,* English trans., London: Warburg Institute, University of London.

Simons, D.J., and Levin, D.T. (1997) "Change blindness", *Trends in Cognitive Science*, in press.

Treisman, A., and Gelade, G. (1980) "A feature-integration theory of attention", *Cognitive Psychology*, 12:97-136.

Treisman, A., and Gormican, S. (1988) "Feature analysis in early vision: Evidence from search asymmetries", *Psychological Review* 95:15-48.

Ullman, S. (1984) "Visual routines", *Cognition* 18:97-159.

Wang, Q., Cavanagh, P., and Green, M. (1994) "Familiarity and pop-out in visual search. *Perception & Psychophysics* 56, 495-500.

Wolfe, J.M., Friedman-Hill, S.R., Stewart, M.I., and O'Connell, K.M. (1992) "The role of categorization in visual search for orientation", *Journal of Experimental Psychology: Human Perception and Performance* 18:34-49.

Wolfe, J.M. (1997) "Visual experience: Less than you think, more than you remember", this volume.

Wolfe, J.M., Yu, K.P., Stewart, M.I., Shorter, A.D., Friedman-Hill, S.R., and Cave, K.R. (1990) "Limits on the parallel guidance of visual search; color X color and orientation X orientation conjunctions", *Journal of Experimental Psychology: Human Perception and Performance* 16:879-892.