2

Attention Routines
and the Architecture of Selection

Patrick Cavanagh

Although we might all know what attention is (James, 1890), we don't all agree on what it does, except that it only does a limited amount of it. This limit, as described by Broadbent (1958), restricts the amount of the available information that can be passed on to higher processing. However, I show that this capacity limit is only one of three independent limits on the information available to higher processing. There is as well an acuity limit: a restriction on the density of items that still permits access to individual items. This resolution of selection is unexpectedly coarse in both space and time. In addition, there is a third bottleneck that I call the coding singularity. I argue that the information available from a selection is neither the raw image detail nor, when there is more than one item in the selection, the set of identities of the items. It is, instead, a single label for the entire pattern. These last two limits, acuity and singularity, may appear to be redescriptions of the same limit but they are not, as I will describe later[1]. The three limits — capacity, acuity, and singularity — are part of a framework for attention that I describe in this chapter. In this framework, attention is not one thing, but a set of routines that are called on to perform specific functions during a task (c.f., Shimamura, 2000; Ullman, 1984). The set of attention routines is identified by one defining characteristic: the initial and final states are reportable but intermediate states are not. I will briefly outline this proposal and then I will focus on the details of one of these functions, selection.

Attention Routines

The idea of attention routines is an extension of Shimon Ullman's proposal of visual routines (1984, 1996). It is also related to Shimamura's (2000) proposals concerning executive attention and working memory where the building blocks of metacognitive control are the functions of selection, maintenance, updating, and rerouting of information. Following the work of Newell and Simon (1972) and others of the early era of cognitive psychology and computer metaphor, Ullman pointed out that many visual tasks could be solved with an explicit, serially executed algorithm. What differentiated Ullman's proposal from those of Newell and Simon and others was that the steps of Ullman's visual routines were not obvious,

nor were they always available to introspection. In the earlier work (e.g., Newell & Simon, 1972), problems like the arithmetic word puzzle DONALD+GERALD=ROBERT (solve for the numerical value of each letter given only that D=5) could be studied by asking subjects to describe their steps out loud as they solved the problem. The subject could report goals, subgoals, and intermediate steps with sufficient detail to allow the experimenters to construct a computer model of the logic of the mental operations involved and infer their reliance on, for example, short term memory and problem space representation.

In the tasks that Ullman examined (e.g., Fig. 2.1), the subject responded rapidly, often within a second or less (Ullman, 1984; Jolicoeur, Ullman, & Mackay, 1986, 1991). The answer appeared with little conscious thought, or with few deliberations that could be reported. Is the X in Figure 2.1 inside or outside the contour? We certainly have to set ourselves to the task but the steps along the way to the answer seem to leave few traces that we can describe explicitly. Unlike the local, hardwired processing of, say, the retina, the processing underlying the analysis of the X and its surrounding contours might well be some serial set of steps that could be captured by an algorithm.

Ullman offered no particular organization for the different types of routines that could be called on and attention was just one of a variety of operators that were available. However, if we organize the routines into three groups, as I have depicted in Figure 2.2, we can remove attention as an agent or resource of any kind and leave it as just the name of a particular level in the hierarchy of routines. This is similar to the way in which, say, a trial by jury is a particular part of the legal system where certain functions take place. We could think of a trial as a resource or an agent but really it is a grouping of functions. In the case of attention routines, we wish to identify what functions these might be and understand how they work. What Ullman had called attention in his 1984 paper now becomes principally one routine, selection, in a larger set that is distinguished by having reportable initial and final states and no reportable intermediate steps.

The hierarchy in Figure 2.2 has three levels: vision routines, attention routines, and cognition routines. They are differentiated by the number of reportable steps each has. Vision routines are inaccessible to awareness. They are implemented outside of conscious control and have no reportable features. Some of these might be hardwired from birth (e.g. computation of opponent color responses), others might emerge with early visual experience (e.g. effectiveness of pictorial cues), and still others may be dependent on extensive practice (e.g. text recognition). Attention routines are consciously initiated by setting a goal or a filter or a
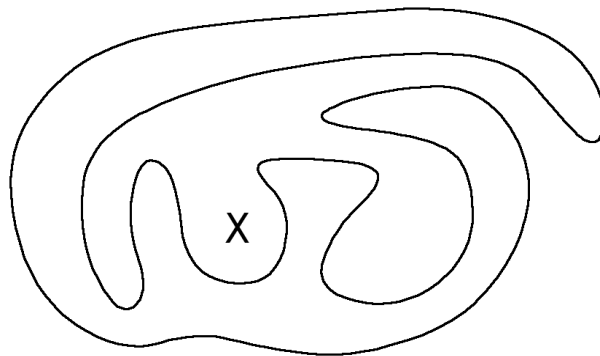

Figure 2.1. Does the X lie within a closed contour? Ullman (1984, 1996) proposed a set of visual routines that would compute the response.

Visual Routines

1. Vision Routines

   Automatic, no introspection

   Examples: Grouping, light constancy, shape recognition, pictorial cues

2. Attention Routines

   Voluntary initiation, reportable output

   No intermediate, accessible states

   Examples: Set selection criteria, spatial & temporal relations, tracking

3. Cognition Routines

   Multiple steps

   Intermediate steps are accessible to introspection

   Examples: Counting, cooking, surgery

Figure 2.2. Three levels of visual routines.

selection target and have a reportable outcome but no reportable intermediate steps. Their intermediate steps are vision routines. Finally, cognition routines are, like the classic cases of Newell and Simon, multi-step routines with several reportable intermediate states, where each individual step is a call to one attention routine.

What have we gained by reshuffling the bits and pieces of a computer metaphor for visual functions? The goal is to produce a list of the specific functions that fall in the attention group and then to find out how these work. This provides an alternative framework for sorting and interpreting the results of the attention literature (c.f., Pashler, 1998; Posner, 1994; Posner & Driver, 1992). Attention routines that begin and end with a reportable state divide the flow of mental activity at its boundaries where the content of awareness changes: new goals are set, new outcomes are computed and these enter and exit awareness as the principal working buffer of these mental tasks. By using reportable states as the syntax to carve up the flow, we identify sets of processes that are bracketed by changes in the content of awareness. For example, if you are asked to report the middle letter of the following three-letter word, ANY, you are aware of the task goals and the word itself to begin with. You also can report the outcome, "n". The initial and final states are reportable but the intermediate states are less accessible. How did you segment the letters, find the middle one and extract it from the word? These steps are highly automated and execute with little or no conscious guidance.

Classically, attention has been characterized by its limited capacity. Here however, that limit is taken to be a consequence of representing the initial and final states of the attention routine in awareness. The suggestion is that the capacity limit is imposed by the constraints of awareness not by attention or attention routines themselves. I will discuss this in more detail at the end of this chapter.

What is a possible list of functions of the attention group? Here is a incomplete, rough set: setting a selection filter, selection itself, binding multiple selections into a compound description, composing descriptions of visual events (Cavanagh, 2003), sending these descriptions to other cortical modules, receiving, decoding and acting on "requests" from other modules (Logan & Zbrodoff, 1999). This is far from even a working list of candidates but it gives the idea of what a list might be like. The purpose of this chapter is not to generate and test routines on this list but to propose that attention is such a group of routines and to pick one particular routine, selection, and describe its function in more detail.

The Architecture of Selection

Attention has been described as the selection of stimuli for higher level processing. I will outline the spatial and temporal constraints on the selection mechanisms and give a description of the type of information that is picked up by selection and what can be done with it. I will start with the constraints on selection and the evidence for multiple, independent, selection operators.

When visual items are selected from a predefined or cued location, many suggest that the effect is like shining a spotlight on that region. Details in the illuminated region are then well defined and can be scrutinized or picked off for further processing. This spotlight metaphor is only a metaphor, of course, but it has been very influential. It has three critical aspects: there is only one "spotlight"; the details within the spotlight can be scrutinized; and selection is limited to this spotlight region. Despite the durable influence of this metaphor, evidence now shows that the first two of its properties, indivisibility and scrutiny, do not hold.

More Than One Spotlight

We know from studies of multiple item tracking (Pylyshyn & Storm, 1987; Yantis, 1992), flanker interference (Driver & Baylis, 1989), and multiple target selection (Awh & Pashler, 2000; Castiello & Umilta, 1992; Kramer & Hahn, 1995) that more than one region can be selected at the same time and that this is not accomplished with simply a larger selection area that encompasses the multiple items. Awh and Pashler (2000), for example, show that the cueing advantage that is seen at both target locations is not seen for targets occasionally presented midway between the two cued locations. These results indicate that there can be more than one selection region. Evidence from multiple item tracking experiments suggests that it may be possible to deploy and control as many as four or five independent selection regions.

How big or small are these selection regions? Many authors have proposed that the selection region can be quite large or quite small depending on the task requirements (e.g., the zoom lens concept, Eriksen & St. James, 1986; Klein & McCormick, 1989; LaBerge, 1983). However, it does appear that there is a minimum size for the selection region and that this smallest size is quite a bit larger than the smallest detail we can see. Our studies of the crowding effect give us some idea of the minimum spacing of objects that still allows selection. If we think of the smallest size of the selection region as the acuity of selection, it varies from 5 times worse than the acuity of vision at the fovea to 30 times worse at only 15 degrees eccentricity (Intriligator & Cavanagh, 2001) and continues this steep downward trend to the limit of the visual field. If our vision were limited to the acuity of selection, we would be legally blind.

Crowding and Acuity of Selection

While looking at the plus sign in Figure 3, you can clearly see the lines to the right but you cannot individuate or count through them, at least not easily beyond the first two. They are within the resolution limit of vision but beyond the limits of selection. This is a demonstration of the crowding phenomenon, more typically explored with letters as in Figure 4.

The traditional view of crowding was that lateral interactions suppress or mask features early in visual processing. This view changed recently when a study in our lab (He, Cavanagh, & Intriligator, 1996) showed that crowded grating patches could induce orientation specific

Figure 2.3. When fixating the plus sign, the vertical bars can be seen. They are thin, parallel, all of equal length. But it is very hard to count them, indicating that at least the middle bars cannot be easily individuated.

aftereffects even though observers were unable to report the orientation of the patches. The stimuli are clearly registered up to at least a level where orientation analysis emerges, level V1 in the visual cortex. Even more recently, Parkes et al. (2001) showed that the orientation of a target in a crowded array of gabor patches could not be reported but, nevertheless, the estimate of the average orientation of the gabors in the array was directly influenced by the orientation of the target. The authors conclude that even though the individual identities in a crowded array are blocked from awareness, they nonetheless do get through to higher levels, albeit in the form of a contribution to the array's overall texture.

These results suggest that the features of a crowded item are not suppressed or masked at an early level. What then accounts for the inability to report the item? We have claimed (He,
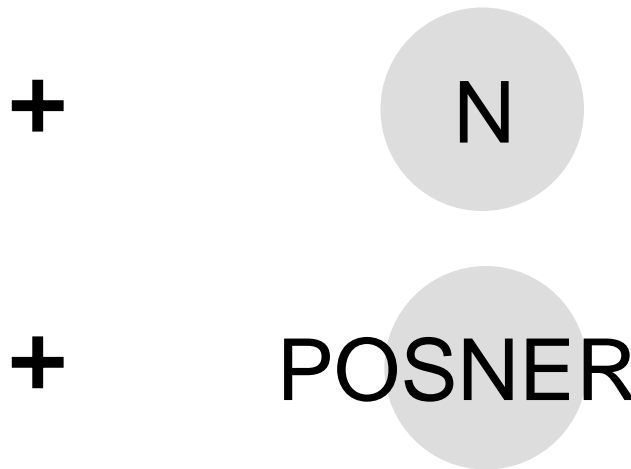

Figure 2.4. When fixating the plus sign on the top left, the N to the right is easily identified. When fixating the plus sign just below, the N seems to meld into a jumbled texture of letters. This is the crowding effect of the surrounding items on the availability of an individual item. Notice that this is not just a function of eccentricity as the letter R, even further in the periphery, can be identified. The N is not in fact lost but contributes efficiently to our judgments of the texture itself (Parkes et al, 2001). The circular gray region depicts a hypothetical, smallest region of selection possible at the eccentricity of the letter N. It enables the selection of the N as an individual letter. When the N is embedded within the string, the same selection region no longer retrieves a single letter but several. If the selection were like a spot light enabling scrutiny within the region of interest, the N would be available as would the S and E. This does not happen. We appear to have access only to a texture-like description within the selection region. This coding singularity prevents access to the raw image and provides only a single description of the entire selection

et al., 1996; Intriligator & Cavanagh, 2001) that this crowding effect is due to limits on the spatial resolution of attentional selection. In severe crowding like that in Figure 4, a target embedded in a dense array of distractors cannot be selected alone, and so cannot be consciously identified or reported. The target is seen only as part of an indivisible, fuzzy grouping that ties together all items within the selection region.

A second experiment showed that the grain of selection is inhomogeneous across the visual field. The task required subjects to mentally step through a set of discs arrayed around a circle while they fixated the center of the circle. Stepping started with one disc that briefly flashed red and then subjects mentally stepped back and forth under computer instruction. After 6 or 7 steps, a probe disc turned red and the subject reported whether it was the disc they had stepped to or not. By varying the number of discs arrayed around the circumference, we measured the density at which the subject could no longer individuate the discs. The results showed that the acuity of selection gets rapidly worse with increasing eccentricity, and is coarser in the upper visual field than in the lower field. Moreover, at locations outside the fovea, acuity is coarser in the radial direction—the direction along the radius from the fovea through the location—than it is in the tangential direction. These were the first measures of the resolution of selection to assess purely the selection of location with no identification required for the target. Our stepping tasks tracked location in dense arrays of identical items and so were unaffected by any preattentive feature interactions that might degrade the recognition measures used in crowding and flanker tasks. The pattern of results for the access to location closely matched those for letter identification in the standard crowding task (e.g., Toet & Levi, 1992). This suggests that the errors of identification in crowding tasks can be attributed to the inability to select the item at the target location. The determining factor is therefore the loss of access to the location and not the loss of the feature information at that location.

Could the resolution of selection depend on attentional load? As a first attempt to address this question we compared the critical radius for crowding with single items (stepping experiment) against the critical radius for crowding estimated from data in the multiple item tracking when four items were being tracked (Intriligator & Cavanagh, 2001). The two estimates matched fairly well. Nevertheless, the critical radius might vary as a function of the number of items being tracked. We addressed this question (Intriligator & Cavanagh, 2001) by varying the number of targets to be tracked from one to four in the multiple item tracking task. Fewer resources are required to track a single target than four targets. As a result, when only a single target is tracked, the surplus resources may be applied to further shrink the selection field around the target. If this is true, there may be no fixed acuity limit for selection, only a size that depends inversely on the resources applied to its maintenance (c.f., Lavie, 1995). In the limit, according to this view, attentional selection for single targets might be limited only by visual resolution.

The experiment showed that this did not happen. Tracking was indeed more accurate at moderately small display sizes when tracking only a single target, but the display size at which tracking became impossible was ultimately the same, no matter how many targets were being tracked. At this point, the moving items could be seen clearly, their spacing was about a factor of ten above the visual resolution limit. Nevertheless, they could not be tracked. This evidence suggests that the spatial limit to selection is an absolute one: it cannot change elastically as more resources become available. It is the same size whether one or four of the regions of selection are being deployed.

In a recent experiment on crowding, we (Tripathy & Cavanagh, 2002) were also able to show that the spatial extent of crowding was fixed, having the same size for large and small targets as well as for luminance-defined and color-defined targets. Local interactions at early

levels in the cortex might be expected to scale with the size of the targets and might also be expected to differ in range for color stimuli and luminance stimuli. Since neither effect was observed, we argue that the spatial extent of crowding is set by the physiological process that accesses the early cortical representation in order to select information. In this case, the size of the region of access would be dependent on the connections from higher levels but independent of target properties.

What is Selected: The Coding Singularity

What can be encoded from each selection region? Clearly, not all features within the region can be picked up as if a piece of the visual image were cut out and passed along for detailed analysis. Many studies have shown that only the features of the target object are selected initially (Egly, Driver, & Rafal, 1994) and not the features of other objects even though they are at the same location (Duncan, 1984). More recently, Blaser, Pylyshyn, and Holcombe (2000) superimposed two gabor patches at the same location and asked subjects to track the features of one or both patches as they smoothly changed their orientations, colors, and bar widths over time. Subjects could keep track of the features of one of the two objects, accurately reporting the direction of any abrupt changes in two properties, say, the target's orientation and color. However, they could not track the properties of both objects at the same location even though they still had to report only the changes in two features, say, the orientation of one and the color of the other.

These results suggest that the information available at a location is not only limited to the features of the target object but also that only one object can be selected at a given location. This remains to be examined in a wider range of situations. Whether or not two objects can be selected at the same location, studies of crowding show that the information available in a single selection is highly constrained.

In the phenomenon of crowding, the inability to identify or even access the location of crowded items implies a fundamental property for the information available for selection. If the selection region were just like a spotlight where details could be scrutinized, a minimum selection region would not create any significant problem. If we could simply select the visual image within the region and scrutinize it (like the gray region in the lower set of letters in Fig. 4), we could read out the letters or items within. We cannot do that.

This result suggests that the information available for selection is some highly processed description or label of the contents of the selection region. If a single, familiar item is within the region, the result of selection is to acquire the label of that item, say, an object like "face", or "shoe", or a letter "B" or number "5" etc. If more than one item lies within the selection region, there may be no description of the combination available from early vision that allows us to then break it down to access each constituent item individually. If two letters are in the selection region, for example, the result might be a description like "letters", or "some letters", or "some lines", depending on what is available. None of these allows us to recover the identities of the letters.

I have used the crowding results to argue for two separate limits: acuity, the minimum spacing of items that allows access to individuals; and singularity, the sparse nature of the description available within the selection. These two limits are closely intertwined. Specifically, if there were no coding singularity, we could scrutinize the details of a selection region and access components down to the limit of visual resolution. The acuity limit of selection would be no coarser than visual acuity. So the acuity limit exists only because of the coding singularity; however, the singularity does not set the value of the acuity limit. The

example of pixellated images makes this point clear. The coding singularity is analogous to the effect of pixellation on a photograph. Within each region, each pixel, there is only one value and no further scrutiny of the photograph is possible at finer scales. The acuity limit of selection described in this chapter is equivalent to the smallest possible size of the pixels. Interestingly, the selection pixels are larger than the available pixels of visual detail. The selection pixels can be scaled up as the regions of interest expand to cover larger areas (as in the zoom-lens models, Eriksen & St. James, 1986; Klein & McCormick, 1989; LaBerge, 1983). However, as these pixels scale up, they do not preserve the new detail they capture, they lose it by giving access to only the single label of the larger area. The singularity limit is equivalent to the claim that the information in one pixel has no internal detail even when the pixel size is scaled up above the minimum size. Selecting a region larger than the minimum size does not provide access to the details of the region at the finest grain but gives a single label for the whole region, no matter how large.

The support for this claim comes from visual search experiments. A number of studies show that it is the highest level descriptions of the items that are initially available and not the underlying features of each item even though the items are not crowded. For example, when searching for an upward curved arc among downward arcs, the search is significantly slower when the arcs are arranged to make mouths and eyebrows of schematic faces (Suzuki & Cavanagh, 1995). We argued that the initial descriptions available for selection are at the face level, not at the constituent feature level. Similarly, He and Nakayama (1992) used test items where one square overlaps another producing an L-shaped region as the only visible part of the square in the back. They showed that subjects could not perform a search for the L shape of the partial occlusion, suggesting that the subjects had access only to the completed square shapes and not the image features. This is also the same point that Hochstein and Ahissar (2002) make in their reverse hierarchy theory. Attention can select initially only at the highest level of description, even when the selection region is much larger than its minimum size. The singularity of coding is not seen only in crowding experiments. It is not a consequence of the small size of the selection region demanded by the crowding paradigm, it is a basic property of selection.

If no scrutiny is available for the information within a single selection, how is compound or complex information selected, bound, and identified. The suggestion here is simply that the four or so independent selection operators pick up local, identifiable pieces and, with some further analysis, construct a combined description (Fig. 5). This suggestion repositions scrutiny and binding from actions operating within a region of selection to actions of an attention routine (bind, group) acting over multiple selections where each individual selection is quite limited in what it can pick up.

Temporal Limits of Selection

The acuity limits of selection in time mirror those of selection in space. In particular, the selection and individuation of events in a rapid stream show very coarse temporal resolution. A spot flickering at slow rates can be perceived as the succession of light and dark phases, but when flickering at a high rate, only flicker is perceived. At even higher rates, the light and dark phases fuse and no flicker is seen. The range of rates over which flicker is visible but the individual light and dark phases are no longer accessible is equivalent to the spatial resolution example in Figure 1 where the spatial pattern of the bars can be clearly seen but individual bars cannot accessed or counted. Many phenomena that involve selection or
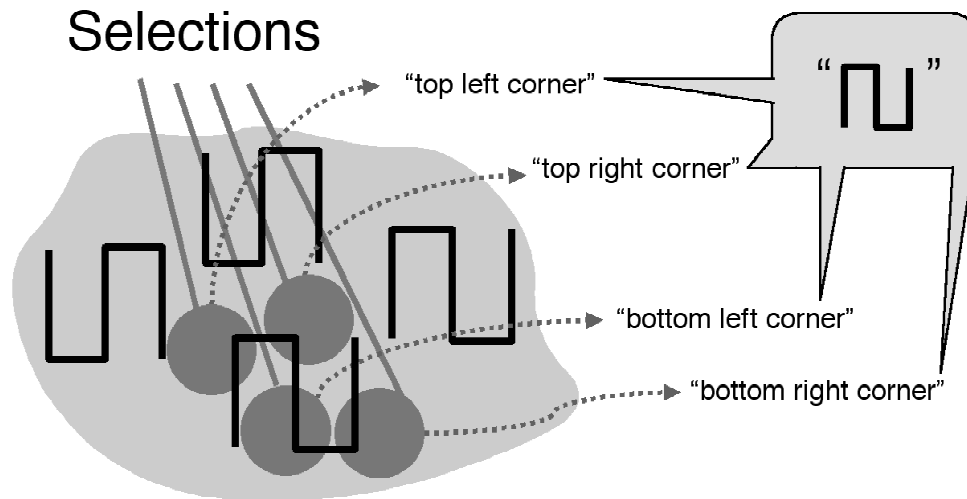
Figure 2.5. Multiple, simultaneous selections can combine the separate parts of an unfamiliar shape. An unfamiliar shape would have no distinctive label or description if it were selected as a whole, other than the equivalent of "some lines". When the four familiar parts (corners) are individually selected, and identified they can then be packaged together with their spatial relations as a distinctive description. This new description can be used to distinguish this shape from, say, the shape to its left.

tracking of sequential events show this low temporal limit of about 4 to 8 Hz (see Verstraten, Cavanagh, & LaBianca, 2000 for a summary). This rate is 8 to 16 frames per second, consistent with the highest rates at which target scenes or words can be identified in Rapid Serial Visual Presentation experiments (Potter, 1993).

Even more important and demanding than selecting disparate bits of a changing scene or object is to integrate and recognize a familiar motion pattern. Many objects and activities have characteristic motions easily revealed in a moment spent imagining the trajectories of butterflies, bike riders, joggers, frisbees, or flapjacks in flight. The recognition of characteristic motions is so robust that a human form is rapidly recognized from the motions of a set of lights attached to a person filmed while walking, running, or performing push-ups in the dark (Johansson, 1973; Neri, Morrone, & Burr, 1998). Although the immediacy of Johansson's point-light walkers suggested that the construction of these motion percepts is effortless, our visual search study (Cavanagh, Labianca, & Thornton, 2001) showed that they are not.

The internal modeling of trajectories and speed can underlie the perception of simple motions as well as complex. A tracked object in smooth (Verstraten et al, 2000) or random motion (Pylyshyn & Storm, 1988) may be supported by the internal representation of its current and expected trajectory. The case of an ambiguous motion (Wertheimer, 1912) is the most compelling example. Wertheimer presented a cross alternating with an X, the 45 deg rotation of the cross. Clockwise or counterclockwise motion could be seen depending on the "set and posture" of attention (Wertheimer, 1912, translated in Shipley, 1961, p. 1070). Only one direction is seen at a time and path seen depends necessarily on the choice of an internal model because the stimulus itself is ambiguous. We have proposed that the high-level motion system (c.f., Anstis, 1980; Braddick, 1980) is anchored by this type of model-based, attention tracking. In other words, it relies on set of attention routines that acquire and track targets (Cavanagh, 1992) and it is limited by the very slow temporal acuity constraints of attention routines (Verstraten, et al., 2000).

Cortical Locus of Selection Mechanisms

Where are the attention routines implemented in the brain? Many studies have suggested that the parietal areas are involved in attentional processing (Corbetta & Shulman, 2002; Gazzaniga & Ladavas, 1987; Posner, Walker, Friedrich, & Rafal, 1984, 1987). Our own imaging and neuropathology studies (Battelli et al., 2001; Battelli, Martini, Barton, & Cavanagh, 2003; Culham et al., 1998; Culham, Cavanagh, & Kanwisher, 2001) indicate that at least the selection routines are strongly represented in the parietal lobes. Our evidence for the parietal locus of selection mechanisms begins with the coarseness and inhomogeneity of the acuity of selection. It is too coarse to be based in cortical area V1 and its variations are more similar to those we would expect for the parietal regions.

The smallest region of selection might be, for example, a hypercolumn since it is not possible to select a smaller unit and represent all the possible features that might occur within the region. We know that there are at least a few thousand hypercolumns in human area V1 but we found evidence for only about 60 independently accessible regions of the central 30° of visual space (Intriligator & Cavanagh, 2001). Moreover, in our study of adaptation and crowding (He et al., 1996), we found that information can be registered in area V1 (as shown by orientation-specific aftereffects) but impossible to select and report. Selection, then, would appear to occur later than area V1.

We also found that the lower visual field has better selection resolution than the upper (He et al., 1996; Intriligator & Cavanagh, 2001). The lower visual field is over-represented in the occipital-parietal regions. This result again suggests that the locus of the selection lies beyond the early retinotopic areas that have relatively similar representations of the upper and lower visual fields. This parietal locus is also supported by our studies of attentive tracking tasks (Culham et al, 1998; Culham, et al., 2001). The areas that are involved in eye movements for tracking and pursuit seem closely linked to those for attentional selection.

Recent results in a study of tracking in our lab (George Alvarez, personal communication, 2003) also suggest that these selection routines may be strongly lateralized in that they operate independently in the left and right hemifields. Tested with the displays shown in Figure 6, the 8 subjects could track 4 items with high accuracy if two of the targets were in
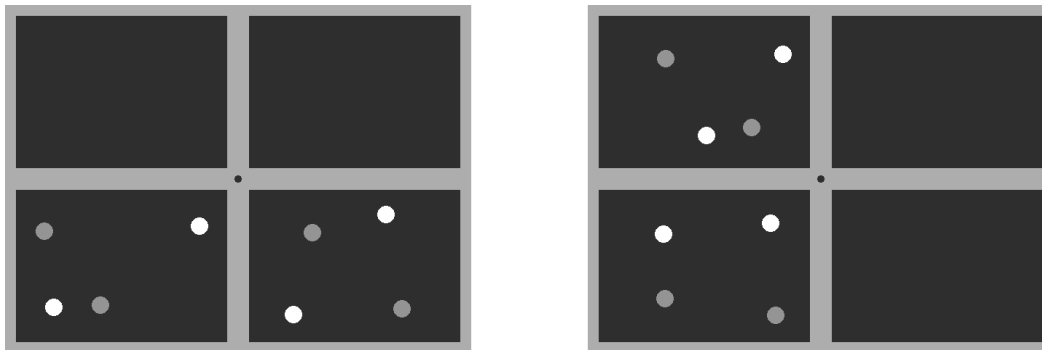

Figure 6. Paired multiple item tracking tasks. On the left, two of the four panels centered at the fixation point have two targets and two distractors each. All items start to move and the white items, which the subject is meant to track, change to gray, indistinguishable from the distractors. On the right is the display where the two tracking tasks are both in the same hemifield. The items bounce off the walls of their respective regions, never crossing the gray separators. Preliminary data (Alvarez, 2003) show that tracking with the two regions in separate hemifields (example on the left) is much easier than when the two regions are in the same hemifield (example on the right).

the left field and two were in the right field. In fact, they were no less accurate tracking two in each field than when tracking just two alone, either in the left or right field. This suggested that the targets in the separate hemifields did not draw on the same capacity. More important, when the four targets were placed in the same field, either left or right, performance plummeted. This implies that the capacity for tracking and the number of independent selection operators is limited within hemifields to about two per field. This result places interesting constraints on the physiological site where selection operators are controlled (it must have independent representations of the two hemifields) and on the process of tracking an object as it crosses the midline where some hand-off may be required between fields. Note that if this result holds up and generalizes to other relevant tasks, the claim of independent tracking capacity in the two hemifields is, in the framework of the attention routines here, a claim about awareness. Specifically, the capacity of awareness must be, to some extent, independently limited for stimuli from the two hemifields. This would not be surprising for split-brain patients, but the suggestion that this holds for some tasks in normals calls for further testing.

Finally, patients who have lesions to the parietal areas show profound deficits in attentive tracking and object enumeration in the hemifield contralateral to the parietal damage (Battelli, et al, 2001). In addition to this contralateral loss in spatial selection, we have recently identified a separate loss in temporal selection. Specifically, patients with right parietal damage have a *bilateral* deficit in differentiating onsets and offsets (Battelli, et al, 2003). In contrast, left parietal patients show no loss in distinguishing onsets and offsets in either field suggesting that this distinction is a function of the right parietal area alone. Since the right parietal patients can detect flickering targets normally, we argued that the deficit occurs at a high level of processing where attention routines sort out which transients belong together. For example, an increase in brightness can be due to the appearance of a bright object or the disappearance of a dark object. The assignment of on and off transients to the appearance and disappearance of objects is contingent on the representation of the objects and their properties. We suggest that this object-based labeling of events is the function of the right parietal lobe. The parietal area is therefore both the site of control for spatial attention routines, where the left and right parietal areas each deal with the contralateral field, and of the temporal aspects of selection, where the right parietal area alone analyzes both hemifields. In this function, the right parietal area could be seen as part of a "When" pathway marking the appearance and disappearance of objects and the beginning and end of events.

Conclusions

The role of attention is now a topic of research in almost all corners of cognitive neuroscience (Fig. 2.7). Despite the recent burst of activity, the basic elements of attention were identified many hundreds of years ago. It was already noted in the 4th century BC that attention appears to have a single focus (Aristotle, trans. 1984), and, in the 1st century BC, that attended items are experienced with more detail and vividness (Lucretius, trans. 1951). Attention could be voluntary (Aristotle, trans. 1984) or involuntary (Augustine, 6th century AD, trans. 1993). In the 11th century AD, Alhazen (trans. 1989) pointed out that some features in vision are apprehended rapidly, in a single glance whereas others require scrutiny. More recently, Wolff (1745) proposed a feature integration theory where attention to the parts of an object is required to integrate them into a whole. A number of these points continue to be debated and refined but the aspects of selection, binding, and limited capacity remain the
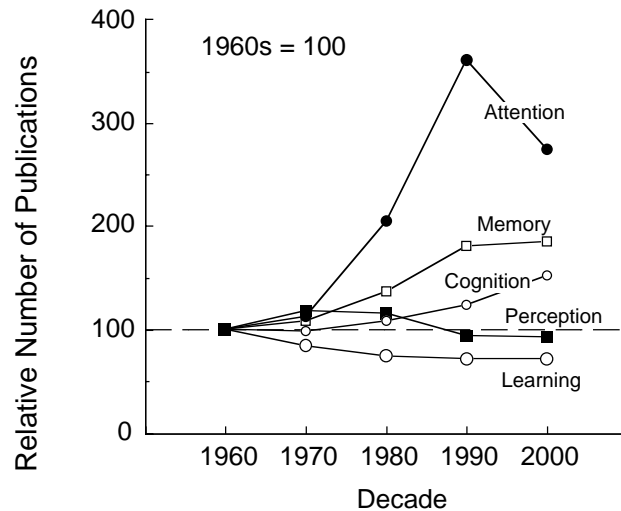
Figure 2.7. Relative numbers of publications for each of five key words, normalized to their relative frequencies in the 1960-1969 decade. The 2000 data span only the years 2000 to 2003.

defining characteristics of attention. Indeed, they are still pretty much all we have (c.f., Hatfield, 1998). As Fodor (2001) said, "That's what so nice about cognitive science, you can drop out for a couple of centuries and not miss a thing" (p. 49).

Despite the development of more sophisticated models (cf. Logan, 1996), there are widely divergent opinions concerning the basic nature of attention. For some authors, attention is an active agent that does things – selects or binds, for example (e.g., Treisman, 1988). For others, it is a resource that must be shared among competing processes (e.g., Kahneman, 1973; Norman & Bobrow, 1975) or a factor that amplifies signals and improves sensitivity (e.g., Lu & Dosher, 1998; Yeshurun & Carrasco, 1998). More recently, some have noted that it is perhaps just a catchall grouping of widely diverse mental functions and phenomena. Driver (2001) has called attention an umbrella term that groups a broad area of research. Walsh (2003) has argued that attention has unrestrained explanatory range; it is freely used for any side of any argument. Like the cosmological ether of pre-20th century physics, says Walsh, it is everywhere, but in the end, nowhere.

The first proposal in this chapter was to divide visual processes according to the availability of reportable states. Attention is then the set of routines that have reportable initial and final states but no accessible intermediate states. Visual routines, the components of these attention routines, have no properties that are accessible to awareness. Cognition routines, processes that comprise sequences of attention routines, have multiple, reportable steps, each one a call to an attention routine. These three levels correspond loosely to the categories of early or preattentive vision (vision routines), to mid- and high-level vision (attention routines), and to cognition or, at least, visual cognition.

The second part of the chapter laid out three processing limits of attention routines. First, there is a capacity limit that is set by the constraints of representing the initial and final states of the routines in awareness. The limit on the number of items simultaneously in awareness appears to be four or five although it may be further constrained as a separate limit of about two for each hemifield. Second, the evidence from crowding experiments suggests the acuity of selection is disturbingly coarse, an order of magnitude or more below spatial and temporal visual acuity. Finally, crowding also revealed a coding singularity. The information available within a single selection is limited to a single, high-level label, affording no opportunity for scrutiny of details within a selection.

One of the two classic characteristics of attention has been its limited capacity. In the definition of attention routines given here, there is no mention of capacity, only of reportability. The initial and final states of an attention routine are consciously registered and reportable. This of course, entails a limit in the ability to simultaneously register multiple items in consciousness. The suggestion is that this is the capacity limit affecting attention. The routines themselves may execute without interference, but when initial states or output values must be posted in awareness, interference arises because of the limited ability to hold multiple items in awareness.

The dependence on awareness and reportability to define attention routines may appear restrictive. We are now comfortable with the notion of reporting our mental states either verbally to others or by motor responses to a computer, or just by introspection to ourselves. But does this criterion put animal research out of reach? I do not believe it does. Alert monkey experiments deal regularly with motor responses that are used to report mental states whether the detection or discriminations of motions, the recognition of target objects, or the enumeration of arbitrary items. These are not reflexes or automated responses although they can become automated with practice.

This very simple criterion raises many questions and I can only note a few here. Reportability, for example, will always underrepresent the actual information that entered and exited awareness. Moreover, some of the limits of attention, attributed here to the capacity of awareness, may be due to limits of working memory, not just limits of visual awareness. The reportability criterion limit might appear to limit this framework to human research. Clearly, many in the alert animal research area allow that responses can be contingent on the mental states of their test animals. In the case of arbitrary responses that are not highly overlearned, the responses should be as valid an index of reportability as a button push is in human psychophysics experiments. However, this view is not without controversy. Finally, the capacity limits of awareness and working memory may not be defined in terms of the number of items that can be held (Luck & Vogel, 1997) but may be more accurately described in terms of the overall amount of detail to be held (for memory, Alvarez & Cavanagh, 2003; for attention, Davis et al., 2001).

The second defining characteristic is that with increased attention, performance improves. This property is of course trivially true in a context of consciously guided task responses. Without selection, there is no awareness of the target and so no explicit response is possible. Any number of implicit measures may reflect the processing of the target by vision routines whose output does not reach awareness but has other effects (e.g. orientation adaptation, He, et al., 1996). Manipulations of selection, like cueing, that affect performance will affect the efficiency of selection in a variety of ways. For example, a non-valid cue may cause selection operators to arrive at the real target location later than they would for a valid cue. Alternatively, the presence of non-valid cues in a session might lead the subject to set one focused selection region at the cued location and one diffuse selection to cover all the remaining locations, reducing the efficiency of selection, and therefore the quality of information retrieved from a miscued target.

Beyond the routines of selection that I have discussed here are many other routines that we would consider part of the attention-based level of processes. Attention routines undoubtedly include feature-based and object-based selection, the packaging and description of selected items, as well as the setting of goals for selection perhaps in response to requests from other modules in the brain. The output of the attention routines must likely be some description of visual events that is shipped off to the other modules of the brain to keep them up to date on the visual world.

Acknowledgments

Note

1. To preview that discussion, clearly there would be no selection acuity limit without the coding singularity, as we could just scrutinize the internal details of a selection down to the limit of visual resolution. However, the existence of a coding singularity does not set the smallest size of the selection region, that smallest size is the acuity limit. Nor is the coding singularity only seen for the smallest selection regions, at the acuity limit in crowded displays. It is revealed on its own in visual search experiments with item spacing much broader than the acuity limit. See page 13 for an expanded discussion.

References

Alhazen. (1989). *The optics of Ibn al-Haytham*. (A. I. Sabra, Trans.) .London: The Warburg Institute.

Alvarez, G. A. & Cavanagh, P. (2003) The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, in press.

Anstis, S. M. (1980). The perception of apparent movement. *Philosophical Transactions of the Royal Society of London B, 290*, 153-168.

Aristotle (1984). Sense and sensibilia. (J. I. Beare, Trans.). In J. Barnes (Ed.), *Complete works of Aristotle*. Princeton: Bollingen.

Augustine (1993). *Confessions*. (F. J. Sheed, Trans.). Indianapolis: Hackett.

Awh E., & Pashler, H. (2000). Evidence for split attentional foci. *J Exp Psychol Hum Percept Perform. 26*, 834-46.

Battelli, L., Cavanagh, P., Intriligator, J., Tramo, M. J., Hénaff, M.-A., Michèl, F., & Barton, J. J. S. (2001) Unilateral right parietal damage leads to bilateral deficit for high-level motion. *Neuron, 32*, 985-995.

Battelli, L., Martini, P., Barton, J. S. S., & Cavanagh, P. (2003). Bilateral deficits of transient visual attention in right parietal patients. *Brain*, in press.

Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature, 408*, 196-9.

Braddick, O.J. (1980). Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society of London B, 290*, 137-151.

Broadbent, D.E. (1958). *Perception and communication*. New York: Pergamon.

Castiello, U., & Umilta, C. (1992) Splitting focal attention. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 837-848.

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

Cavanagh, P. (2003) The language of vision. The *Perception* Lecture. *Perception, 32*, Suppl,. 1.

Cavanagh, P., Labianca, A. T., & Thornton, I. M. (2001). Attention-based visual routines: Sprites. *Cognition, 80*, 47-60.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci., 3*, 201-15.

Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. H. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology, 80*, 2657-2670.

Culham, J. C., Cavanagh, P., & Kanwisher, N. G. (2001) Attention response functions: characterizing brain areas with fmri activation during parametric variations of attentional load. *Neuron, 32*, 737-745.

Davis, G., Welch, V. L., Holmes, A., & Shepherd A. (2001). Can attention select only a fixed number of objects at a time? *Perception, 30*, 1227-48.

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology, 92*, 53-78.

Driver, J., & Baylis, G.C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception & Performance, 15*, 448-456.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*, 501-17.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123*, 161-177.

Eriksen, C.W., & St. James, J.D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics, 40*, 225-240.

Fodor, J. (2001). *The mind doesn't work that way.* Cambridge: MIT Press.

Gazzaniga, M.S., & Ladavas, E. (1987). Disturbances in spatial attention following lesion or disconnection of the right parietal lobe. In M. Jeannerod (Ed.), *Neurophysiological and neuropsychological aspects of spatial neglect, Vol. 45* (pp. 203-213). Amsterdam, Netherlands: Elsevier Science Publishing Co, Inc.

Hatfield, G. (1998). Attention in early scientific psychology. In Richard D. Wright (Ed.). *Visual attention. Vancouver studies in cognitive science, Vol. 8.* (pp. 3-25).

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature, 383*, 334-337.

He, Z. J., & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature, 359*, 231-233.

Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron, 36*, 791-804.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43*, 171-216.

James, W. (1890). *The principles of psychology.* New York: Holt.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*, 201-211.

Jolicoeur, P., Ullman, S., & Mackay, M. (1986). Curve tracing: a possible basic operation in the perception of spatial relations. *Mem Cognit., 14*, 129-40.

Jolicoeur, P., Ullman, S., & Mackay, M. (1991). Visual curve tracing properties. *J Exp Psychol Hum Percept Perform., 17*, 997-1022.

Kahneman, D. *Attention and Effort.* Englewood Cliffs, NJ: Prentice-Hall, 1973.

Klein, R., & McCormick, P. (1989). Covert visual orienting: Hemifield-activation can be mimicked by zoom lens and midlocation placement strategies. *Acta Psychologica, 70*, 235-250.

Kramer, A.F., & Hahn, S. (1995). Splitting the beam: Distribution of attention over noncontiguous regions of the visual field. *Psychological Science, 6*, 381-386.

LaBerge, D. (1983). Spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception & Performance, 9*, 371-379.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *J Exp Psychol Hum Percept Perform. 21*, 451-68.

Logan, G. D. (1996). The CODE theory of visual attention: an integration of space-based and object-based attention. *Psychol Rev., 103*, 603-49.

Logan, G. D., & Zbrodoff, N. J. (1999). Selection for cognition: Cognitive constraints on visual spatial attention. *Visual Cognition, 6, 55-81*

Lu, Z. - L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research, 38*, 1183-1198.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. Nature, *390*, 279-81.

Lucretius (1951). *On the nature of the universe.* (R. Latham, Trans.). London: Penguin Books.

Nakayama, K. (1990). The iconic bottleneck and the tenuous link between early processing and perception. In Colin Blakemore (Ed.), Vision: Coding and efficiency. (pp. 411-422). Cambridge: Cambridge University Press.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton.

Neri, P., Morrone M. C., & Burr, D. C. . (1998). Seeing biological motion. *Nature, 395*, 894-896.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, N.J.: Prentice-Hall.

Norman, D. A. & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7*, 44-64.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nat Neurosci., 4,* 739-44.

Pashler, H.(1998). *The psychology of attention.* Cambridge, MA: MIT Press.

Posner, M. I. (1994). Attention: the mechanisms of consciousness. *Proc Natl Acad Sci U S A., 91,* 7398-403

Posner, M. I., & Driver, J. (1992). The neurobiology of selective attention. *Curr Opin Neurobiol., 2,*165-9.

Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal injury on covert orienting of attention. *J Neurosci., 4,* 1863-74.

Posner, M.I., Walker, J.A., Friedrich, F.A., & Rafal, R.D. (1987). How do the parietal lobes direct covert attention? Special Issue: Selective visual attention. *Neuropsychologia, 25,* 135-145.

Potter, M. C. (1993). Very short-term conceptual memory. *Mem Cognit., 21,* 156-61.

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

Sharma, V., Levi, D. M., & Klein, S. A. (2000). Undercounting features and missing features: evidence for a high-level deficit in strabismic amblyopia. *Nature Neuroscience, 3,* 496-501.

Shimamura, A. P. (2000). Toward a cognitive neuroscience of metacognition. *Consciousness and Cognition, 9,* 313–323.

Shipley, T., Ed. (1961). *Classics in Psychology.* New York: Philosophical Library.

Suzuki, S., & Cavanagh, P. (1995). Facial organization blocks access to low-level features: an object inferiority effect. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 901-913.

Toet, A., & Levi, D.M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research, 32,* 1349-1357.

Treisman, A. (1988). Features and objects: the fourteenth Bartlett memorial lecture. *Q J Exp Psychol A., 40,* 201-37.

Tripathy, S. P., & Cavanagh, P. (2002). The extent of crowding in peripheral vision does not scale with target size. *Vision Research, 42,* 2357.

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

Ullman, S. (1996). *High-level vision.* Cambridge, MA: MIT Press.

Verstraten, F. A.J., Cavanagh, P. & Labianca, A. T. (2000). Limits of attentive tracking reveal temporal properties of attention. *Vision Research, 40,* 3651-64.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Science,* in press.

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie, 61,* 161-165.

Wolff, C. (1745). *Psychologie ou traité sur l'âme.* Reprinted 1998, Hildesheim: Georg Olms Verlag.

Yantis, S. (1992). Multi-element visual tracking: attention and perceptual organization. *Cognitive Psychology, 3,* 295-340.

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature, 396,* 72-75.