

The Role of Neural Mechanisms of Attention in Solving the Binding Problem

Review

John H. Reynolds* and Robert Desimone
Laboratory of Neuropsychology
National Institute of Mental Health
Bethesda, Maryland 20892

For purposes of this review, we will define the binding problem as the problem of how the visual system correctly links up all the different features of complex objects. For example, when viewing a person seated in a blue car, one effortlessly sees that the person's nose belongs to his face and not to the car, and that the car, but not the nose, is blue. To fully understand the solution to this problem requires a good neurobiological theory of object recognition, which does not exist. We will therefore follow the lead of the computer engineer, who, when asked to describe how he would write a computer program to recognize a chicken, replied, "first, assume a spherical chicken." Thus, in this review we will make some assumptions that simplify the binding problem in order to appreciate how neural mechanisms of attention provide a partial solution.

Evidence for the Binding Problem

Before getting to the mechanisms of attention, it is useful to consider the psychological and physiological evidence that the binding problem even exists. The classic psychological evidence comes from studies of "illusory conjunctions" (Treisman and Schmidt, 1982; Cohen and Ivry, 1989; Ivry and Prinzmetal, 1991; Arguin et al., 1994; Prinzmetal et al., 1995). In a typical experiment, human subjects are briefly presented with an array containing several different objects, such as letters of the alphabet, shown in different colors. In one condition, subjects are cued to attend to one of the letters, and in a comparison condition their attention is divided between the array and another object. In the former condition, with undivided attention, the letter is perceived correctly. However, in the latter condition, with divided attention, subjects often perceive the wrong combinations of letters and colors, e.g., a red letter B and green letter C are misperceived as a green B and a red C. That is, color and shape are incorrectly bound.

These studies argue that the binding problem exists and that attention helps to solve it. Consistent with this interpretation, damage to the parietal lobes, which are thought to be involved in allocating attention, can result in illusory conjunctions during free viewing. Patient R. M. suffered two successive strokes that caused extensive bilateral parietooccipital lesions that spared the temporal and frontal lobes. He was severely impaired in attentionally demanding visual tasks, and, when presented with two colored letters, he often misconjoined their identities and colors, even after viewing for up to 10 s (Friedman-Hill et al., 1995). By contrast, he was much

less impaired on less attentionally demanding tasks such as detecting a "popout" stimulus.

The physiological evidence for the binding problem comes from studies of neurons in extrastriate visual cortex of primates. One key observation is that different features of an object are processed to a certain extent by different neurons within the visual system. For example, some neurons in areas V2 and V4 respond selectively to the orientation of an object, independent of its color, whereas other V2 and V4 neurons respond selectively to color, independent of shape, and cells selective for different features are often located in separate regions of both areas (Schein et al., 1982; Livingstone and Hubel, 1983; Desimone et al., 1985; Tanaka et al., 1986b; Schein and Desimone, 1990; Ghose and Ts'o, 1997). Still other neurons in the middle temporal area (MT) and the medial superior temporal area (MST) encode various aspects of the motion of the stimulus (Desimone and Ungerleider, 1986; Saito et al., 1986; Tanaka et al., 1986a, 1989; Rodman and Albright, 1989; Stoner and Albright, 1992; Geesaman and Andersen, 1996; Treue and Andersen, 1996; Bradley et al., 1998; Buračas et al., 1998; O'Keefe and Movshon, 1998; Lisberger and Movshon, 1999). Logically, in order to identify the shape, color, and motion of a stimulus, the visual system must somehow integrate the activity of these different shape-selective, color-selective, and motion-selective neurons. When only one stimulus is present in the visual field, this is not a difficult problem because these features can only be assigned to one possible stimulus. However, when multiple stimuli appear together in the visual field, which is the typical situation in "real-world" scenes, the visual system must assign the correct color, shape, and motion signals to each object.

If all neurons were highly selective for spatial location, then illusory conjunctions might be avoided by integrating features separately at each precise location in the visual field. For example, one might integrate color and shape only for neurons that share the same receptive field. However, as one moves through the ventral visual stream that underlies object recognition (Ungerleider and Mishkin, 1982), the receptive field size of neurons increases steadily. Neurons in area V4, for example, typically have receptive fields that are several degrees wide near the representation of the center of gaze, and neurons in the inferior temporal (IT) cortex have receptive fields that can include the entire central visual field, on both sides of the vertical midline (Desimone et al., 1984; Desimone and Schein, 1987; Gattass et al., 1988). These large receptive fields are presumably necessary to recognize large complex objects and may mediate the ability to recognize objects of any size as the same, regardless of their retinal location. The ventral stream cortical areas with large receptive fields are known to play an important role in object recognition, because recognition is greatly impaired when these areas are damaged (Cowey and Gross, 1970; Gross et al., 1971; Dean, 1976; Ungerleider and Mishkin, 1982).

Similarly, information about the motion of visual patterns appears to be processed in areas within the dorsal

* To whom correspondence should be addressed (e-mail: reynolds@ln.nimh.nih.gov).

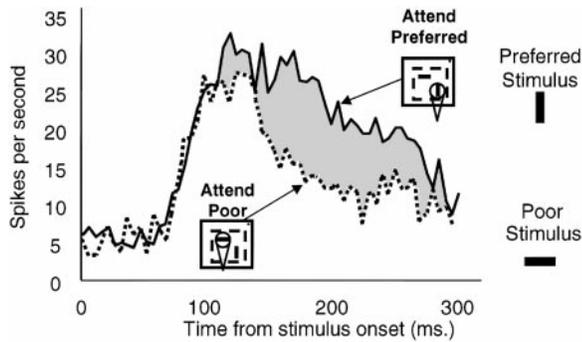


Figure 1. Average Response over Time of a Population of Stimulus-Selective V4 Neurons to Two Stimuli inside the Classical Receptive Field

When attention was directed to the poor stimulus (bottom line), responses were lower than when attention was directed to the preferred stimulus (top line).

stream that have large receptive fields. When a complex pattern, such as a plaid, moves through the visual field, neurons in primary visual cortex respond selectively to the direction of movement perpendicular to the lines that make up the plaid. However, some neurons in area MT, an area with large receptive fields, respond according to the direction of motion of the plaid as a whole (Gizzi et al., 1983; Movshon et al., 1985; Rodman and Albright, 1989; Stoner and Albright, 1992). Position-invariant object identity and global pattern motion are just two of many examples of stimulus attributes that appear to be represented in areas whose neurons have large receptive fields.

The number of potentially erroneous feature conjunctions increases exponentially with the number of objects in a large receptive field (Tsotsos, 1990). As receptive fields become larger and larger at each processing stage of the ventral stream, there is therefore an increasing number of erroneous feature bindings to rule out. Thus, the binding problem emerges as a necessary consequence of the large receptive fields found in higher-order areas.

Attention and Shrinking Receptive Fields

A possible solution to the binding problem was suggested by the study of Moran and Desimone (1985), who found that when two stimuli appear within the receptive field of a neuron in either area V4 or inferior temporal cortex, the response elicited by the pair depends on which of the two stimuli is attended. They chose the shape and color of the stimuli such that one of the stimuli elicited a strong response when it was presented alone (the *preferred* stimulus), whereas the other elicited a very weak response when it was presented alone (the *poor* stimulus). When attention was directed to the preferred stimulus, the pair elicited a strong response. However, when attention was directed to the poor stimulus, the identical pair elicited a weak response, even though the preferred stimulus was still in its original location (see Figure 1).

Luck et al. (1997a) and Reynolds et al. (1999) replicated this result in area V4 and found it to hold in area V2 as well. Treue and Maunsell (1996) have reported the

same pattern of results in areas MT and MST. They presented a preferred stimulus (a dot moving in the cell's preferred direction) together with a poor stimulus (a dot simultaneously moving in the opposite direction) within the receptive field. The responses elicited by the pair were higher when attention was directed to the dot moving in the preferred direction, relative to when attention was directed to the dot moving in the opposite direction. A similar experiment recently conducted by Seidemann and Newsome (1999) also found that when two stimuli appear within the receptive field of an MT neuron, attention to the more-preferred stimulus increases responses, relative to when attention is directed to the poorer stimulus. However, the magnitude of these effects was much smaller than that observed by Treue and Maunsell (1996), suggesting that the degree to which attention is able to modulate neuronal responses and the stage of processing at which this occurs may be task dependent.

One way to account for these results is to assume that when attention is directed to one of two stimuli within a cell's receptive field, this causes the receptive field to constrict around the attended stimulus, leaving the unattended stimulus outside the receptive field (see Figure 2). According to this interpretation, when attention is directed to the preferred stimulus, the neuron is driven by the preferred stimulus, and its response is therefore large. When attention is directed to the poor stimulus, the preferred stimulus is now excluded from the receptive field, the cell is driven by the poor stimulus, and its response is small. Thus, according to this interpretation, attention solves the binding problem by increasing the effective spatial resolution of the visual system so that even neurons with multiple stimuli inside their large receptive fields process information only about stimuli at the attended location.

Further support for the idea that attention increases the spatial resolution of the visual system comes from two recent psychophysical studies. In one (Yeshurun and Carrasco, 1999), subjects were tested in three different tasks that required them to make fine spatial discriminations. In all three tasks, subjects responded more slowly and less accurately when the stimulus appeared at more peripheral locations, where receptive fields are larger. And, in all three tasks, directing attention to the location of the stimulus resulted in faster and more accurate performance. In a related study, Yeshurun and Carrasco (1998) also found that attention paradoxically impairs performance in a task that requires processing of low-spatial frequency components of a stimulus. Subjects were asked to detect the presence of a texture-defined target, which required integration of information at low spatial resolution. Unlike most visual tasks, performance on this detection task is poorer at the fovea than it is at mid-peripheral locations, where spatial resolution is most appropriate for the task (DeValois and DeValois, 1988; Graham, 1989). Yeshurun and Carrasco found that attention improved performance on this task at peripheral locations, presumably by increasing the spatial resolution of peripheral vision to better fit the task. Strikingly, attention significantly reduced performance on the foveal task, where further improvement of spatial resolution would be expected to undermine performance.

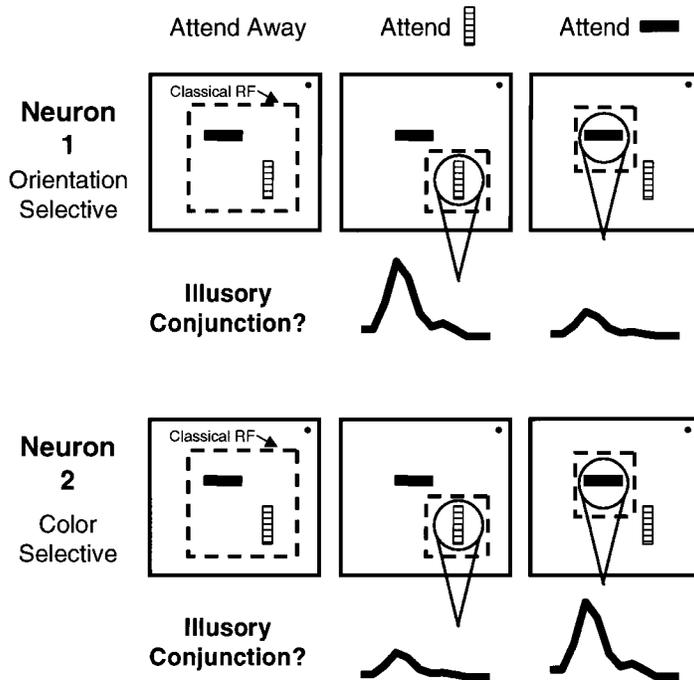


Figure 2. Schematic Diagram Illustrating How Attention May Resolve Illusory Conjunctions. Each row of three panels shows a schematic representation of the responses of a single hypothetical neuron. The orientation-selective neuron appearing in the upper row responds to a vertical bar appearing anywhere in its classical receptive field (dashed box at left), regardless of its color. The neuron appearing below responds selectively to a dark bar, regardless of its orientation. When attention is directed away from the receptive field, and both stimuli appear together, both neurons will be somewhat active, indicating the presence but not the exact location of their preferred features. Because the precise location of the two stimuli is not represented in the largely spatially invariant responses of the neurons, the neuronal response cannot unambiguously indicate whether the two features belong to the same stimulus. When attention is directed to one of the stimuli, both neurons respond according to the features of the attended stimulus. Since neurons respond as though only the attended stimulus were present, their activities can be unambiguously interpreted as reflecting features of a common, attended object.

The Role of Competition

Desimone and Duncan (1995) have proposed that such changes in spatial resolution may emerge as a result of competitive interactions between stimuli. According to this hypothesis, multiple stimuli in the visual field activate populations of neurons that engage in competitive interactions, possibly mediated through local, intracortical connections. When subjects are instructed to attend (or choose voluntarily to attend) to a stimulus at a particular location or with a particular feature, this generates signals within areas outside visual cortex. These signals are then fed back to extrastriate areas, where they bias the competition in these areas in favor of neurons that respond to the features or location of the attended stimulus. As a result, neurons that respond to the attended stimulus remain active while suppressing neurons that respond to the ignored stimuli. In other words, neuronal responses are now determined by the attended stimulus, and any unattended stimuli are filtered out of their classical receptive fields—an effective increase in the neurons' spatial selectivity.

For example, imagine recording from a neuron that responds vigorously to stimulus A and fails to respond to stimulus B. If attention is directed to stimulus A, this will bias the competition in favor of the population of cells that normally responds to A, and the cell being recorded will remain active. If attention is then directed to stimulus B, the competing population will win, and the cell being recorded will be suppressed, along with the other members of its population. In retinotopically organized areas, such as area V4, this competition is thought to be strongest for cells located near to one another in the cortex, which therefore share similar receptive fields.

We recently tested this idea that attention works through competitive processes by recording V2 and V4

neuronal responses in a behavioral paradigm that allowed us to isolate automatic sensory processing mechanisms from attentional ones (Reynolds et al., 1999). We first tested cells for competitive interactions in the absence of attention. While the monkey attended to a location far outside the receptive field of the neuron, we measured the response to a single reference stimulus within the receptive field. We then compared this response to the response when a probe stimulus was added within the receptive field. When the probe was added to the field, the neuron's response was drawn toward the response that would have been elicited if the probe had appeared alone. For example, the response to a preferred reference stimulus was typically suppressed when a poor stimulus was added as a probe, even when the poor stimulus elicited small excitatory responses when it appeared alone. Symmetrically, the response of the cell increased when a preferred probe stimulus was added to a poor reference stimulus. Thus, the response of a cell to two stimuli in its field is not the sum of its responses to both but rather is a weighted average of its response to each alone.

To test how attention influenced this automatic competitive mechanism, we then had the monkey attend to the reference stimulus. The effect of attending to the reference stimulus was to almost precisely eliminate the excitatory or suppressive effect of the probe. If, in the absence of attention, the probe stimulus had suppressed the response to the reference, then attending to the reference restored the cell's response to the level that had been elicited when the reference was presented alone (Figure 3A). Conversely, if the probe stimulus had increased the cell's response, attending to the reference stimulus drove the response down to the level that had been recorded when the reference was presented alone (Figure 3B). Thus, the effect of attention was to modulate the underlying competitive interaction between stimuli.

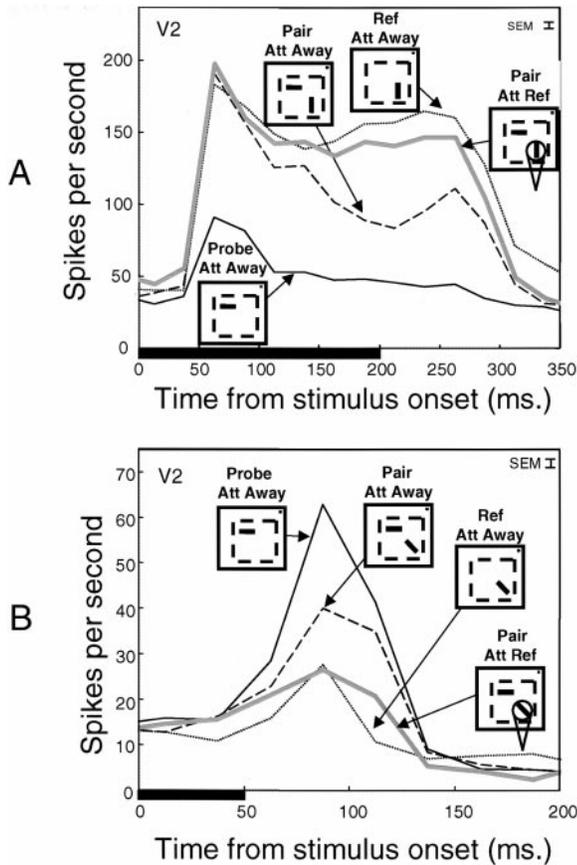


Figure 3. Responses of Single Neurons in Area V2, Showing that the Attended Stimulus Controls the Response of the Neuron (A) The dotted line at the top shows the response of a single V2 neuron to a stimulus of its preferred orientation and color. The solid line at the bottom shows the response to a poor stimulus. The dashed line in the middle shows the response when the two stimuli are presented together, with attention directed away from the receptive field. The addition of the poor stimulus strongly suppresses the response elicited by the preferred stimulus. When attention is directed to the preferred stimulus, this suppression is eliminated, and the neuron responds as though the poor stimulus were absent. (B) Here, the response of another V2 neuron to a poor stimulus (bottom dotted line) is increased (middle dashed line) by the addition of a preferred horizontal stimulus. When attention is directed to the poor stimulus, the cell's response is reduced to a level that is comparable to its response to the poor stimulus, alone.

Given the close similarities between attention effects in the dorsal and ventral streams, it seems likely that both streams use the same underlying competitive circuit. Consistent with this possibility, studies of neuronal responses to multiple stimuli have found opponent direction suppression in areas MT and MST of the dorsal stream (Qian and Andersen, 1994; Recanzone et al., 1997; see also Mikami et al., 1986). Responses to stimuli moving in a nonpreferred direction are increased by the addition of a second stimulus moving in the preferred direction in the receptive field, and, conversely, responses to preferred stimuli are suppressed by the addition of a stimulus moving in the null direction. It remains for future studies to confirm that attention modulates this underlying competitive circuit.

Recent functional magnetic resonance imaging (fMRI)

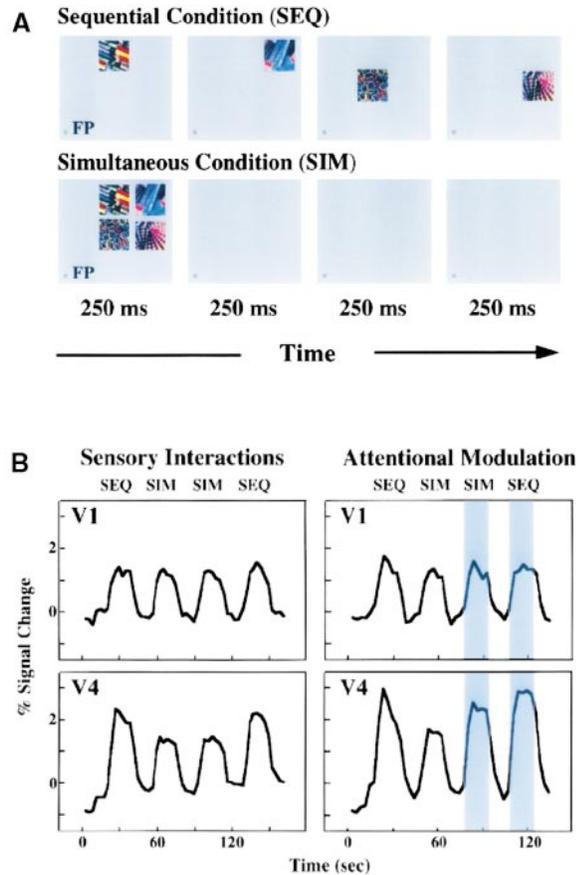


Figure 4. fMRI Signals in the Human Show Sensory Interactions and Attention Effects that Scale with the Size of Receptive Fields at Successive Cortical Areas

(A) Task. Subjects fixated a spot while stimuli appeared either asynchronously (top four images) or simultaneously (bottom four images). The total amount of time each stimulus appeared was constant. Subjects either performed an attentionally demanding task or attended to one of the four stimuli.

(B) fMRI signals. In area V1 (top images), where receptive fields are too small to include more than one stimulus, the presence of additional stimuli did not suppress neuronal responses, and attention had very little effect. However, in area V4 (bottom images), whose receptive fields are large enough to include multiple stimuli, presenting the stimuli simultaneously caused a suppressed response. This is illustrated in the reduced magnitude of the two middle peaks (labeled "SIM SIM") in the lower left image. When attention was directed to one of the stimuli (lower right image, time period labeled in blue), this suppression was largely filtered out.

experiments in humans support the idea of attentional modulation of an underlying sensory competition. Kastner et al. (1998) compared the average fMRI signal elicited by a stimulus when it was presented alone versus the signal elicited when the same stimulus appeared simultaneously with other stimuli (Figure 4A). They found that when attention was directed away to another location, the presence of the additional stimuli reduced the strength of the fMRI signal in the human analogs of monkey V4 and TEO, areas whose receptive fields are large enough to encompass multiple stimuli (Figure 4B). This suppressive effect was minimized when stimuli were separated from one another in space, consistent with the idea that competition is greatest for stimuli

occupying the same receptive field. Suppression was weakest in primary visual cortex, whose smaller receptive fields would rarely be expected to process information from more than one of the stimuli, resulting in minimal interactions between stimuli. When subjects were instructed to attend to stimuli at one of the locations, this eliminated most of the suppressive effect of the distractor stimuli, consistent with the physiological results in monkeys.

Evidence for the Attentional Bias

These experiments provide support for the idea that illusory conjunctions are resolved when attentional feedback signals bias an underlying competitive circuit, causing neurons to respond exclusively to the attended stimulus. Possibly the most direct physiological evidence for this predicted bias was reported by Luck et al. (1997a), who found that attending to a position within the receptive field of a neuron increased its spontaneous firing rate. The spontaneous activity of V2 and V4 neurons increased by 30%–40% when attention was directed to the location of the receptive field, even when the field contained no stimulus. A recent fMRI study in humans has found similar evidence for a sustained increase in activity with attention in the absence of visual stimulation, with the increases occurring at the retinotopic locus of the attended location in the visual field (Kastner et al., 1999). This effect was found in several cortical visual areas of the dorsal and ventral streams. In the physiological study of Luck et al. (1997a), the shift in firing rate was largest when the monkey attended to the “hot spot” of the receptive field, where stimuli elicited the strongest response, and was smallest when attention was directed toward the edge of the field. Thus, consistent with its putative role in selecting out one of several stimuli from within the receptive field, the attentional signal that gives rise to this change in spontaneous firing rate has a higher spatial resolution than the receptive field.

The bias in favor of an attended stimulus or location is also evidenced by an increase in response to a stimulus at an attended location, which has been found in many, but not all, physiological studies. With just a single stimulus inside the receptive field, Spitzer et al. (1988) found that when animals were required to perform an attentionally demanding task using that stimulus, responses of neurons in area V4 increased by 18%, compared to when the monkey's attention was directed outside the field. More recently, Connor et al. (1996) have reported that V4 responses to stimuli appearing within the receptive field are sometimes enhanced when the monkey attends to a location just outside the receptive field. McAdams and Maunsell (1999) have found V4 responses to be increased by 26% when monkeys discriminated the orientation of a counterphase grating within the receptive field, as compared to the response elicited by the same stimulus when the monkey discriminated the color of a stimulus appearing outside the receptive field (Figure 5A).

Not all studies have found consistent enhancement of responses with attention to a single stimulus inside the receptive field, however. Moran and Desimone (1985), Haenny et al. (1988), Maunsell et al. (1991), Motter

(1993), and Luck et al. (1997a) have found either no effect of attention on the response to a single stimulus or changes in response that were nearly equally likely to be increases as decreases.

A possible explanation for this puzzle has recently been reported by Reynolds et al. (1996, *Soc. Neurosci.*, abstract; 1997, *Assoc. Res. Vis. Ophthalmol.*, abstract), who found that V4 neurons have increased contrast sensitivity when attention is directed to a position within the receptive field (Figure 5B). Neurons responded to faint attended stimuli that failed to elicit a response when attention was directed away from the receptive field. On average, attention caused cells to respond as though the luminance contrast of the stimulus had been increased by 22%. This increased sensitivity resulted in large increases in the neuronal response to low-contrast stimuli. The effect of attention to a single stimulus inside the receptive field was greatly reduced for high-contrast stimuli, which nearly saturated the neuronal response even in the absence of attention. Nicholas et al. (1996, *Soc. Neurosci.*, abstract) found that this increase in sensitivity was also reflected in enhanced responses to texture-defined stimuli. As with luminance contrast, the effect of attention to a single stimulus was diminished for texture-defined stimuli that were well segregated from their backgrounds. Together, the results indicate that although the bias in favor of an attended stimulus can easily be detected with stimuli that are low contrast, the effect of the bias may be missed if the response of the cells has already been saturated by highly salient stimuli that are easily discriminated from their backgrounds. Thus, variations in stimulus contrast, in addition to variations in task and task difficulty, may help explain the variability across studies in the effects of attention to single stimuli. One possible way of conceptualizing the effect of attention on single stimuli is illustrated in Figure 5C. Orientation tuning is largely invariant to changes in contrast (Skottun et al., 1987). For stimuli that are below saturation contrast, an increase in physical contrast will result in a gain multiplication of the orientation tuning curve. Similarly, a leftward shift of the contrast–response function resulting from increased neuronal sensitivity with attention also predicts a gain multiplication of the orientation tuning curve, which has been reported by McAdams and Maunsell (Figure 5C). Although they did not test stimuli across a range of contrasts in that study, the contrast used may have been below the saturation level of the cells.

Nonspatial Attention

In the cases considered above, attention was directed to a given location, and the features of the object at that location could be bound together by filtering unattended objects from the receptive fields of extrastriate neurons. How can attention solve the binding problem when the location of the object of interest is unknown in advance, e.g., when searching for a face in a crowd? In a recent study, Chelazzi et al. (1998) found that when monkeys search for an object in an array based on its features, this can modulate receptive fields of neurons in inferior temporal cortex in a manner similar to what happens in spatially directed attention (see Figure 6). In this task, monkeys viewed a cue stimulus, followed by a brief

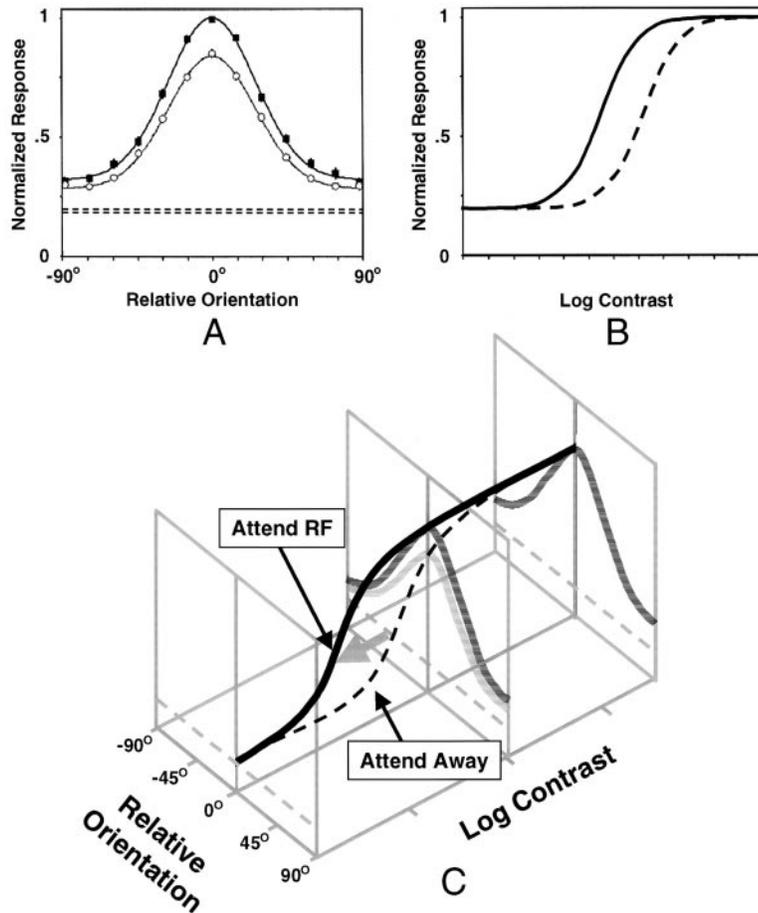


Figure 5. Gain Multiplication and Increased Sensitivity with Attention in Area V4

(A) McAdams and Maunsell (1999) found that attention caused a multiplicative increase in the orientation tuning curve.

(B) Reynolds et al (1996, Soc. Neurosci., abstract; 1997, Assoc. Res. Vis. Ophthalmol., abstract) found that attention caused an increase in V4 neurons' contrast sensitivity, as reflected in a leftward shift of the contrast-response function.

(C) A possible way to reconcile the two findings. The change in response resulting from changes in orientation and contrast are orthogonal. Thus, an increase in contrast sensitivity (leftward shift in the contrast-response function) will result in a multiplicative increase in the tuning curve.

delay, after which an array of stimuli (the *search array*) appeared within the neuron's receptive field. The monkeys' task was to indicate whether or not a target matching the cue stimulus was present in the array. The positions of the stimuli in the search array were selected at random, so the monkey could not know where the target would appear, if it appeared at all. During the delay period, many inferior temporal neurons had a higher baseline firing rate when the cue stimulus was a preferred stimulus for the cell than when it was a poor stimulus. This is analogous to the increase in baseline firing rates found in other studies when spatial attention was directed to a cell's receptive field (Luck et al., 1997a), and it is evidence for a bias in favor of neurons representing the target stimulus. Within 150–200 ms after the search array appeared, the neuron's response increased or decreased, depending on whether the cue was, respectively, a preferred or a poor stimulus for the cell. That is, shortly after the onset of the search array, the cell responded primarily to the target stimuli and responses to nontarget stimuli were suppressed. Thus, attention to an object feature apparently has effects on competitive interactions that are similar to those observed with attention to a spatial location.

Likely Sources of the Biasing Signal

Prefrontal and parietal cortex have a number of properties that implicate them as possible sources of task-dependant biasing signals in posterior visual areas. Pre-

frontal cortex shows elevated spontaneous responses that encode remembered spatial locations and remembered objects (Wilson et al., 1993; Rainer et al., 1998). It also has anatomical connections by which it could transmit these signals to appropriate dorsal and ventral stream areas (Barbas, 1988; Barbas and Pandya, 1989; Ungerleider et al., 1989). Parietal cortex also encodes remembered spatial locations in a form that could be read out as a spatial biasing signal and has the requisite anatomical connections to do so. Stimuli appearing at an attended location are preferentially represented in parietal cortex (Bushnell et al., 1981; Gottlieb et al., 1998), and different parts of parietal cortex provide a variety of different spatial reference frames that would be helpful in directing attentional feedback signals to behaviorally relevant locations in space (reviewed by Andersen, 1997; Colby and Goldberg, 1999). Patients with bilateral parietal lesions have difficulty binding together object features, despite the fact that these features are known to be processed primarily outside of parietal cortex. This would be expected from a lesion that eliminates an important source of signals that bias competitive interactions within the cortical areas that encode these features. Consistent with this, recent fMRI studies (Corbetta et al., 1993, 1995; Nobre et al., 1997; Kastner et al., 1999) have also found the superior parietal lobule in the human to be highly active during a spatial attention task.

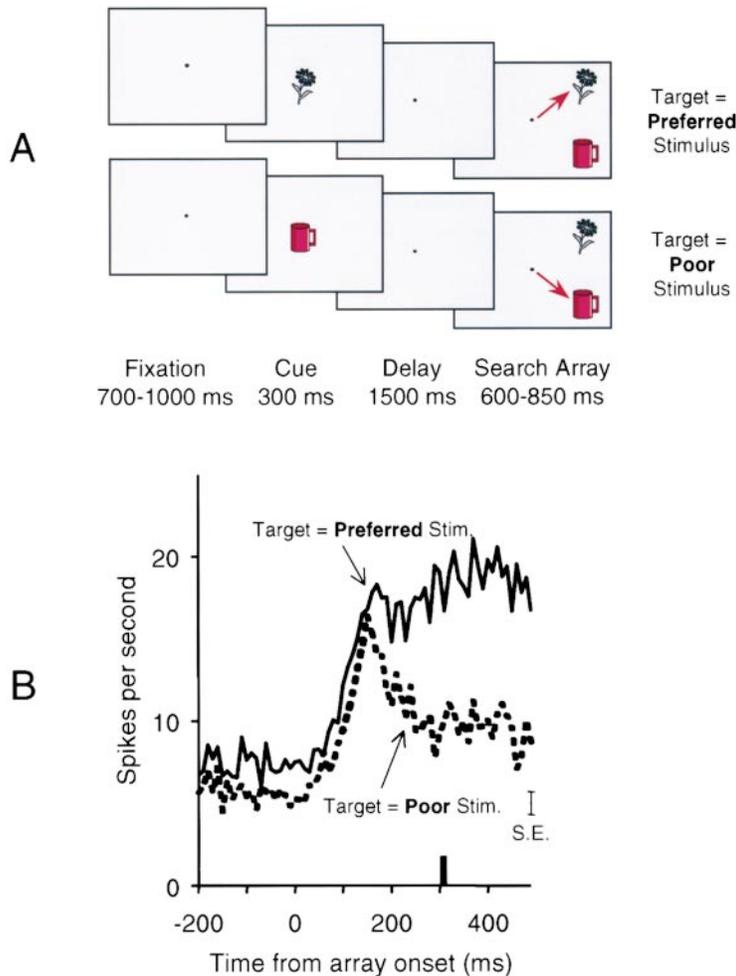


Figure 6. Responses of IT Neurons during Memory-Guided Search

(A) Task. Monkeys fixated a spot on a computer screen and were shown a central cue (here, either the flower or the cup). After a delay, two (or more) stimuli appeared within the receptive field, and the monkey had to saccade to the stimulus that had appeared earlier as the cue. Sometimes (top four images), the cue was a preferred stimulus for the cell (the flower) and the monkey had to saccade to the flower. On separate trials (lower four images), the cue was a poor stimulus (the cup) and the monkey had to saccade to the cup.

(B) Neuronal responses. During the delay period, IT neurons showed an elevated baseline activity that reflected the cue stored in memory. The spontaneous rate was higher on trials in which the cue was a preferred stimulus for the cell, relative to trials when the cue was a poor stimulus. After the search array appeared, the response separated, increasing or decreasing depending on whether the cue was, respectively, a preferred or poor stimulus for the cell. This separation occurred well before the onset of the saccade, which is indicated by the vertical bar on the horizontal axis.

Stimulus Salience Helps Ameliorate the Binding Problem

The picture that emerges from these considerations is that attentional feedback from areas such as prefrontal and parietal cortex eliminates illusory conjunctions by biasing competition in favor of stimuli appearing at the attended location or in favor of the searched-for object. The finding that attention causes extrastriate neurons to respond as though the salience of the stimulus had been increased (Nicholas et al., 1996, *Soc. Neurosci.*, abstract; Reynolds et al., 1996, *Soc. Neurosci.*, abstract; 1997, *Assoc. Res. Vis. Ophthalmol.*, abstract; Gottlieb et al., 1998) raises the intriguing possibility that physical salience itself may also play a part in resolving the binding problem. If high salience biases the same competitive mechanisms as does attention, this should cause neurons to "lock on to" whatever stimulus is most salient. This could partially ameliorate the binding problem for unattended stimuli by setting the visual system to bind together the features of the most salient stimulus by default. That is, when attention is not actively engaged in processing a stimulus within the scene, or when attention is directed away from the receptive field, the most salient element in the scene might dominate neuronal responses.

A recent experiment of Reynolds and Desimone (1997,

Soc. Neurosci., abstract) supports this hypothesis. We reasoned that if salience acts as a bottom-up bias on the same competitive mechanisms that appear to subserve attentional selection, then neuronal responses to a pair of stimuli should be dominated by the most salient stimulus within the receptive field. To test this, two stimuli were presented within the receptive fields of V4 neurons, while the monkey attended away to a location outside the receptive field. We varied the relative contrast of the two stimuli and compared the response elicited by the pair to the responses elicited by the two stimuli presented individually. The neuron's response to the pair was similar to its response to the higher-contrast stimulus presented alone. That is, the cell responded as though the low-contrast stimulus were not present. This result was most striking for highly selective cells, where increasing the contrast of the poorer stimulus significantly reduced the response to the pair, despite the fact that the poor stimulus alone elicited a small excitatory response. Consistent with the hypothesis that attention and high salience bias the same competitive mechanisms, the two effects were additive. When the monkey directed its attention to the higher-contrast stimulus, the neuron's response was independent of the presence of the lower-contrast, unattended stimulus. Thus, normal variations of salience in the visual environment may

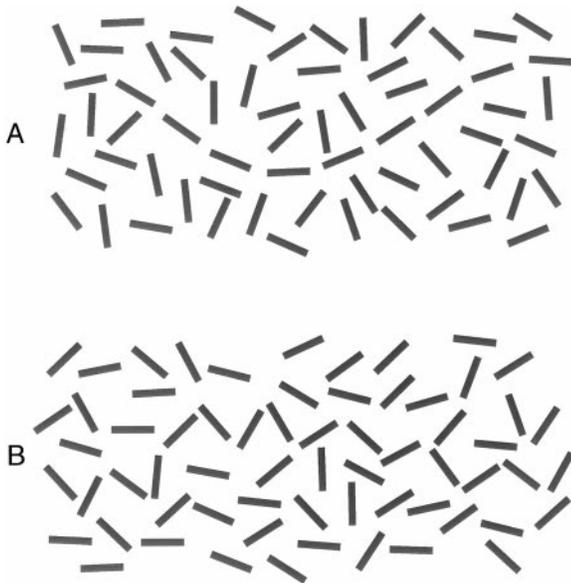


Figure 7. Saliency Can Emerge from Grouping
(A) A contour of nearly collinear line elements pops out of a background composed of randomly oriented line elements.
(B) If the elements connecting end-to-end are not roughly collinear, they do not form a salient grouping.

serve to partially ameliorate the binding problem even for unattended objects, by transferring control over neuronal responses to the most salient elements of the scene.

Clearly, this is only a partial solution to the general binding problem. Almost all physiological studies to date have used well-isolated, simple stimuli that could be segregated out from the background on the basis of changes in local luminance contrast. This leaves unresolved the difficult question of what actually competes for processing—is it local luminous flux, higher-order object features, or entire integrated “objects”? This difficult aspect of the binding problem remains unresolved, but studies of perceptual grouping suggest one way in which grouping principles could play a role. When individual elements form a perceptual group, they gain saliency and stand out from their background, as illustrated in Figure 7. In the upper panel, line elements that are roughly collinear cohere into a contour that stands out from the background, like a snake in the grass. The lower panel also includes a set of elements that connect end-to-end, but this grouping does not pop out because the elements are not collinear.

Kapadia et al. (1995) found an analogous effect of grouping on the responses of neurons in striate cortex. When a line element within the receptive field is presented together with additional, collinear elements outside the receptive field, the responses of many striate neurons are enhanced. Changing the spatial relationships between elements in ways that break up this perceptual group (such as introducing an intervening perpendicular element, separating the elements, or making them noncollinear) reduces or eliminates this enhancement. Saliency from perceptual grouping, represented by increased responses of neurons in striate cortex, is

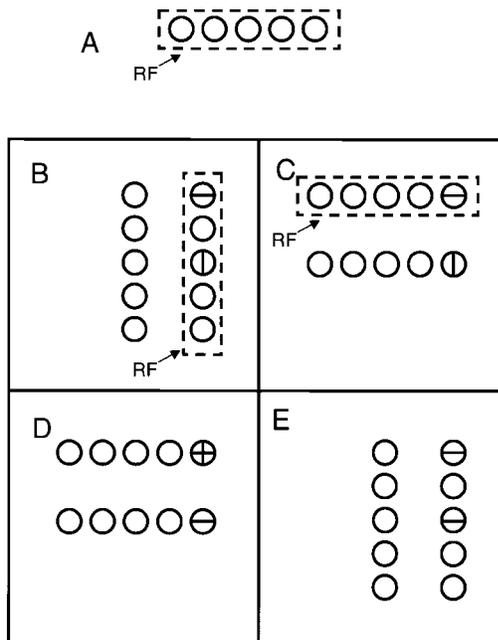


Figure 8. A Speculative Hypothesis about Illusory Conjunctions and Perceptual Grouping

(A) Suppose that when elements form a perceptual group, such as this row of stimuli, receptive fields shrink around the group.
(B) According to this hypothesis, the vertical and horizontal line elements fall within common receptive fields. As a result, neuronal response will be invariant across positions within the group, and the features may be misconjoined to form a plus.
(C) When the two elements appear within separate groups, we assume that they will not fall within a common receptive field and will not be misconjoined. Prinzmetal (1981) found that illusory conjunctions are more likely among elements within a group than among elements in separate groups.
(D and E) Control conditions in which illusory conjunctions cannot occur because the plus is actually present (D) or because the two features were of the same orientation (E).

likely to influence higher-order extrastriate neurons in much the same way as saliency arising directly from high luminance contrast and texture popout. It remains to be seen whether perceptual groups defined by factors such as collinearity, closure, similarity, symmetry, proximity, common motion, and common onset will turn out to have enhanced control over neuronal responses in higher-order areas with large receptive fields.

This is clearly a speculative hypothesis, but it does make an interesting prediction. We have presented evidence that illusory conjunctions emerge as a result of the spatial uncertainty that results from the large receptive fields found in extrastriate cortex. If, as a result of the more salient representation of groups in earlier cortical areas, higher-order extrastriate receptive fields effectively shrink around perceptual groups (see Figure 8A), this would separate elements from different groups into separate receptive fields. Illusory conjunctions should therefore be more likely between features appearing within a perceptual group than between equally spaced features appearing in separate perceptual groups.

This is exactly what was found in a study conducted by Prinzmetal (1981). Arrays of circles appeared briefly,

followed by a mask. They were either arranged to form two columns (as in Figure 8B) or two rows (as in Figure 8C). The arrays typically included a vertical and a horizontal line segment, either appearing at the same location, to form a plus (as in Figure 8D), or appearing at separate locations (as in Figures 8B and 8C). The two line segments could either appear within the same perceptual group (as in Figure 8B) or in separate groups (as in Figure 8C). Prinzmetal measured how often subjects misconjoined the vertical and horizontal line segments to form an illusory plus. In a comparison condition (see Figure 8E), the two line segments were of the same orientation, so they could not form an illusory conjunction. Subjects often incorrectly reported seeing a plus symbol when the line segments were perpendicular to one another. Despite the fact that the physical separation of the two line segments was identical, subjects were significantly more likely to form illusory conjunctions between elements within a perceptual group than when they appeared within separate groups.

As described earlier, sensory interactions and attention effects are greatest when two stimuli appear within a common receptive field. This provides a way to probe the hypothesis that receptive fields shrink around salient perceptual groups. It will be interesting to see whether competitive sensory interactions and associated attentional modulations will turn out to be greater among elements within a perceptual group than among elements in separate groups.

Biological Constraints on Models

The results reviewed above provide several constraints on models of the circuitry underlying attentive visual processing, and hence provide insight into some of the neural mechanisms involved in solving the binding problem. First, with attention directed away from the receptive field, the effect of adding a second stimulus inside the receptive field is to draw the neuron's response toward the response elicited by the added stimulus alone. If we change the identity of the second stimulus so that it elicits a smaller response when presented alone, it typically becomes proportionally more suppressive.

Second, neuronal responses are biased toward the more salient stimulus. If a preferred stimulus is paired with a poor excitatory stimulus, the suppression caused by adding the poor stimulus typically increases with the contrast of the poor stimulus.

Third, when attention is directed to a location within the receptive field, this increases the neuron's sensitivity to stimuli appearing at that location, as reflected in the cell's ability to respond to low-salience stimuli that do not elicit a response when unattended.

Fourth, when multiple stimuli appear together, attending to one of them causes the neuronal response to be biased toward the response that would have been elicited if the attended stimulus had appeared alone. Thus, attending to the poorer of two stimuli typically reduces the neuronal response to the pair.

Fifth, the control bias arising from differences in relative salience adds to the bias arising from attentional

feedback. Directing attention to a higher-contrast stimulus causes it to exert even greater control over the neuronal response, while attending to the lower-contrast stimulus counteracts the relative contrast bias. This enables the neuron to respond to an attended low-contrast stimulus, even in the presence of a high-contrast distractor.

Sixth, attention has the capacity to modulate the responses of neurons at a spatial scale that is smaller than the scale of an individual receptive field. When multiple stimuli appear within the receptive field, attention causes the neuron to selectively process one stimulus within the receptive field while filtering out others. This high spatial resolution is also reflected in the attentionally induced increases in spontaneous firing rate, which vary in magnitude as a function of precisely where in the receptive field attention is directed.

Seventh, the effect of attention depends on spatial separation of stimuli. Two stimuli can be placed far enough apart that they will fall within separate receptive fields in one cortical area, say V2, while remaining close enough together to fall within a common receptive field at a subsequent area. Here, the effect of attention changes from simple enhancement to a shift of control, as illustrated in Figure 9. Thus, at one stage, attention serves to boost the strength of the stimulus-evoked response, separating it from one form of noise: spontaneous activity. At the subsequent stage, attention filters out a second type of noise: the signal that is elicited by the unattended stimulus.

A Model Based on Biased Competition

One possible model that satisfies these constraints is illustrated in Figure 10. The model assumes that when two stimuli appear within the visual field, they will, at some stage of cortical processing, activate separate populations of neurons. The two circles at the bottom of the figure represent these two input populations. The circuit is assumed to be repeated at each stage of the cortical hierarchy, but to fix ideas, let us suppose these input populations are in area V2. If the stimuli appear close enough to one another, they will activate input populations that project to a common output neuron in the subsequent stage of cortical processing (say, V4). The circle at the top of the figure is intended to represent this output neuron. Each input population is assumed to provide both excitatory and inhibitory drive to the output neuron, through inhibitory interneurons, which are not shown in the figure. The response of the output neuron is assumed to depend on the ratio of excitatory to inhibitory drive from the input populations. The selectivity of the output neuron for the two stimuli therefore depends on the strength of excitatory and inhibitory inputs projecting from each input population. The stimulus whose input population provides proportionally more excitatory input will elicit a relatively stronger response in the output neuron.

The model accounts for the observed relationships between selectivity, sensory interactions, and attention effects, as follows. When the two stimuli appear together, their excitatory inputs are assumed to be additive, as are their inhibitory inputs. Therefore, when a second stimulus is added to the receptive field, the output neuron's response will shift toward the response

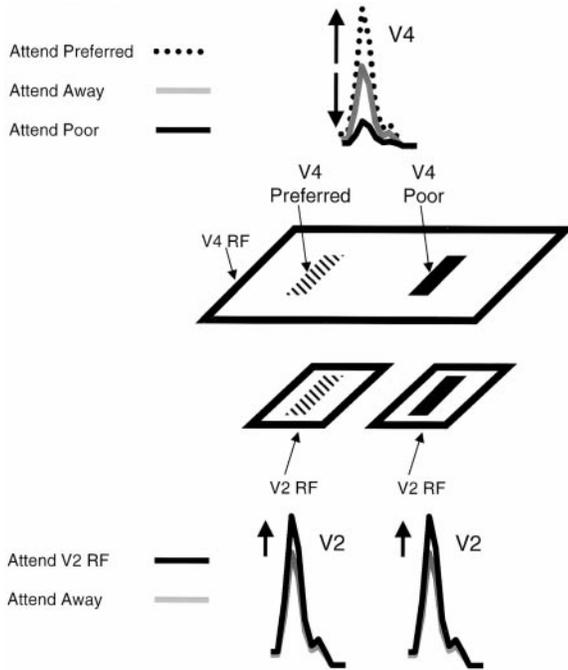


Figure 9. Dependence of Attentional Modulation on Receptive Field Size

A qualitative change in the effect of attention takes place when information passes from a cortical area with receptive fields too small to include both stimuli to a successive cortical area with receptive fields that are large enough to encompass both stimuli. At the bottom of the figure, we show the schematic responses of two neurons in an area with small receptive fields. The hatched stimulus on the left falls within the receptive field of the neuron whose response is illustrated on the left. The solid stimulus on the right falls within the receptive field of the other neuron, on the right. For both of these hypothetical neurons, the response is magnified when attention is directed toward the stimulus within the receptive field. The larger box above illustrates the larger receptive field of a higher-order neuron. Here, the two stimuli fall within a common receptive field. When attention is directed to the cell's preferred stimulus (the hatched bar), this causes an increase in response. However, when attention is directed to the poor stimulus, this results in a reduced response.

elicited by the second stimulus alone, as observed experimentally. If the physical salience of one of the stimuli is increased, this is assumed to increase the magnitude of the response elicited in the input population, and the output neuron's response will shift toward the response that would be elicited by the more salient stimulus alone. Attention is assumed to act in an analogous manner, by increasing the efficacy of synapses projecting from the neuronal population activated by the attended stimulus. Thus, attention will increase contrast sensitivity at the attended location. This change of synaptic efficacy will also increase the output neuron's sensitivity to spontaneous activity of afferent neurons, resulting in a baseline shift. The spatial selectivity of this baseline shift follows from the spatial specificity of the projections from afferent neurons to the output neuron. The model accounts for the finding that attention effects are minimal for high-contrast stimuli, because at high contrast the response of the output neuron is already saturated by strong excitatory and inhibitory inputs, and further increases of input strength will not cause further increases in response.

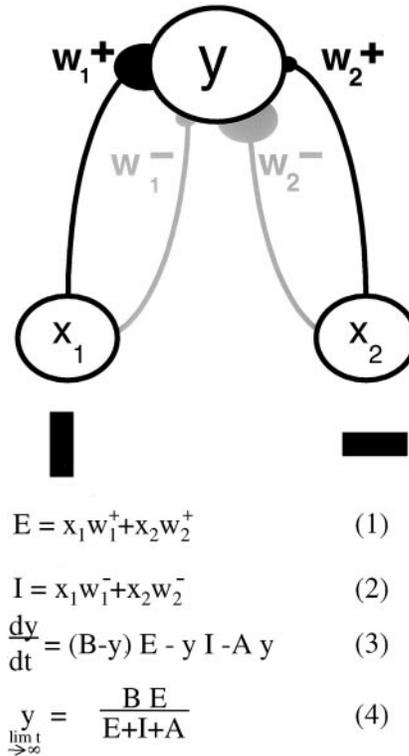


Figure 10. One Possible Implementation of the Biased Competition Model

The circle on top represents the neuron being recorded, whose firing rate is designated by the variable y . The two circles below represent populations of "input" neurons that respond to the reference and probe stimuli, respectively, and that project to the upper neuron. The average responses of the input populations are designated x_1 and x_2 . Black lines represent the excitatory projections from each input population to the measured cell, and gray lines indicate the inhibitory projections, which are assumed to depend on inhibitory interneurons (not shown in figure). The variables w_1^+ and w_2^+ stand for the magnitudes, or weights, of the excitatory projections from the two input populations, while w_1^- and w_2^- stand for the weights of the inhibitory projections. Equations 1 and 2 state, respectively, that the total excitatory and inhibitory inputs to the cell are the responses of the input populations, weighted, respectively, by their excitatory and inhibitory weights. Equation 3 describes the change in the response of the output neuron over time. Equation 4 describes the equilibrium response of the output neuron, which depends on the total mix of excitatory and inhibitory input. The model is further described in Reynolds et al. (1999).

The effect of attention is unmasked by the presence of a second stimulus inside the receptive field because attention must now filter out the large signals elicited by the competing stimulus. Increasing the efficacy of inputs from the attended population causes the mix of excitatory and inhibitory inputs, and thus the output neuron response, to be dominated by the attended stimulus. Input strength can be adjusted by either a change in synaptic efficacy or a change in the strength of response in the input population, so the effects of attention and relative contrast are additive. The high spatial resolution of attentional modulation arises because the input neurons, whose synapses are assumed to be modulated by attention, have the requisite fine spatial resolution. When the two stimuli appear together so that they both

appear within the receptive field, a change in the efficacy of synapses projecting from one input population will filter the nonattended stimulus out of the V4 receptive field.

The biased competition model provides a unified framework within which to think about attention and its role in resolving the binding problem. However, the implementation sketched above is only one of a number of possible models that satisfy the constraints derived from the neurophysiological literature on attention. Among the existing alternatives are models that implement competitive interactions using lateral inhibitory connections and assume that the attentional bias is mediated by a direct excitatory signal or by triggering synchronous discharge among cells whose receptive fields overlap with the focus of attention (see, e.g., Koch and Ullman, 1985; Anderson and Van Essen, 1987; Niebur et al., 1993; Olshausen et al., 1993; Ferrera and Lisberger, 1995; Grossberg, 1995, 1999a, 1999b; Stemmler et al., 1995; Pouget and Sejnowski, 1997; Borisjuk et al., 1998). Additional experiments will be necessary to further constrain the set of possible models.

Relationship between Our Proposal and Other Theories of Illusory Conjunctions

The explanation we have proposed for the role of attention in resolving illusory conjunctions is related to the Feature Integration Theory (FIT) proposed by Treisman and Gelade (1980) (see also Treisman and Schmidt, 1982; Treisman, 1996). Like FIT, our explanation depends on the proposition that illusory conjunctions arise from the decomposition of visual stimuli into their constituent features and that attention is necessary to avoid illusory conjunctions of the features.

Our proposal differs from FIT in several ways. First, we propose that illusory conjunctions arise specifically because of the spatial uncertainty that results from the large receptive fields found in higher-order visual areas, and that attention resolves this uncertainty by biasing competitive cortical mechanisms in favor of the attended stimulus. This proposal is similar in spirit to the Ambiguity Resolution Theory proposed by Luck et al. (1997b). In addition, our proposal is that attentional feedback is only one of many factors that can bias competition in favor of a particular object. We predict that other factors, such as variation in the relative salience of stimuli, which also bias competition, can reduce or eliminate illusory conjunctions. As demonstrated by Chelazzi et al. (1998), working memory can also bias competitive interactions in a way analogous to spatial attention. This could be crucial in avoiding illusory conjunctions between overlapping objects, where spatial attention alone cannot, in principle, link features together correctly. Our model offers a way to explain how a spatially distributed perceptual group of elements may be linked together into a whole and segregated out from its background, by making use of the greater salience of grouped elements to put the grouped elements at a competitive advantage over ungrouped elements. Finally, because our model is implemented at the neuronal level, it offers a way to bridge the gap between the psychophysics and neurophysiology of attention.

Conclusions

Findings from neurophysiology, psychophysics, and fMRI all converge on a common conclusion about the role of attention in solving one aspect of the binding problem: illusory conjunctions. When multiple unattended stimuli appear within the receptive field of an extrastriate neuron, the neuronal response cannot unambiguously be associated with any one stimulus, and it is this ambiguity that leads to illusory conjunctions. Attentional mechanisms eliminate illusory conjunctions by filtering out unattended stimuli whose features could be misconjoined with those of the attended stimulus.

This selection process occurs in several stages and depends on attentionally induced increases in the effective salience of the attended object in earlier stages of processing, where it appears alone within a receptive field. As signals from multiple stimuli progress forward into higher-order areas with larger receptive fields, stimuli compete to control neuronal responses. The added strength of the signals from the attended stimulus resolves this competition in its favor. As a result, the responses of higher-order neurons with large receptive fields encode only the attended stimulus, implicitly binding together its features. In the absence of attentional control, a highly salient stimulus (e.g., based on higher luminance or perceptual grouping principles) may also bias the same competitive mechanisms, helping to resolve illusory conjunctions among unattended objects. Finally, all of these mechanisms for attentional selection and salience operate in concert within multiple cortical areas of dorsal and ventral visual processing streams, implicitly binding together the features encoded by different neurons throughout extrastriate cortex.

Acknowledgments

We thank L. Chelazzi, V. Ferrera, P. Fries, S. Kastner, and A. Rossi for helpful comments on the manuscript.

References

A comprehensive reference list for all reviews can be found on pages 111–125.