Specialized Representations in Visual Cortex: A Role for Binding?

Geoffrey M. Ghose and John Maunsell
Division of Neuroscience and Medicine
Howard Hughes Medical Institute
Baylor College of Medicine
Houston, Texas 77030

Seeing is deceptively simple. We perceive objects, symbols, movements, and other aspects of the visual scene without effort or awareness of the mechanisms that process visual information. Our continuous and seamless visual perceptions depend on the activity of billions of individual neurons. The first step in seeing an object is the generation of a pattern of activity that is distributed across hundreds, if not millions, of photoreceptors. These in turn activate many other neurons in more central structures.

How the activity of widely distributed neurons can lead to unitary percepts has been a key question in neuroscience dating to the origins of the modern concept of neurons. The apparent continuity of perception was one of the major philosophical arguments against Cajal’s neuron doctrine. How could discrete anatomical units be responsible? It has been an enduring subject of inquiry since then. For example, Köhler and Held (1949) suggested that unitary visual perception depends on currents that flow through the cerebral cortex as if it were a volume conductor, a concept which led to experiments that tested perceptual capability after embedding wires or insulators into cortex to disrupt the hypothesized fields (Lashley et al., 1951; Sperry et al., 1955).

An important component of this issue is binding: the process of linking together the attributes of a perceptual object. A typical scene contains many distinct objects, any one of which might be behaviorally relevant to the observer. Each object will activate a widely distributed population of neurons responsive to its constituent attributes. These neurons may be interspersed with or separated by neurons that represent other attributes. Any process that links together the activity of neurons that represent a specific object is said to bind its representation.

Several distinct forms of neuronal processing could be called binding. The elaboration of receptive field properties, such as the center-surround receptive field of a retinal ganglion cell, can be considered a type of binding because it links together the activity of related neuronal signals. Neuronal activity must also be bound over time to give continuity to stimuli; otherwise, it would be impossible to keep track of moving objects or interpret any extended pattern of sensory input, such as speech. Both of these types of binding must occur in the nervous system.

Our discussion is restricted to the temporal correlation hypothesis (Singer and Gray, 1995), which addresses a different and specific type of binding: the labeling of groups of simultaneously active neurons that represent elements of a given object. It proposes that 30–70 Hz modulation of neuronal activity synchronizes the firing of neurons that represent elements of a given object, and that this correlated firing labels the neuronal activity associated with one object. Different frequencies or phases of modulation could be used to simultaneously label different objects. With this approach, neurons distributed across visual cortex, or across different sensory and motor systems (Engel et al., 1997), could be dynamically and selectively bound whenever their activity was associated with a single object.

The temporal correlation hypothesis is based on many neurophysiological studies that have shown that the spike rate of cortical neurons sometimes oscillates at frequencies between 30 and 70 Hz, and that neurons driven by a single stimulus sometimes oscillate in synchrony (reviewed by Engel et al., 1992b, 1997; Singer and Gray, 1995). This synchronization has been observed between neurons in different visual areas (Eckhorn et al., 1988; Engel et al., 1991b; Roelfsema et al., 1997), between sites in different cerebral hemispheres (Engel et al., 1991a; Nowak et al., 1995b), and between sensory and motor regions (Bressler et al., 1993; Murthy and Fetz, 1996a, 1996b).

The temporal correlation hypothesis is limited. It does not address how a particular group of neurons is segmented from other active cells, or how synchrony is achieved. Nor does it attempt to explain how synchrony is measured and acted upon to form unitary perceptions, decisions, or actions. Its potential value lies in describing an intermediate mechanism in the process of getting from distributed patterns of neuronal activity to unitary decisions and actions.

The temporal correlation hypothesis proposes a specific mechanism to bind distributed representations, but whether such binding occurs has not been established. In the sections that follow, we consider whether this type of binding is a necessary step for visual processing. We suggest that there is no compelling need for binding by temporal correlation, and that visual performance could instead depend on the existence of small groups of neurons with highly specialized response properties. We then review data on neuronal specialization in visual cortex and argue that they are consistent with this alternative view. Even if neurons have highly specialized response properties, any stimulus will activate many neurons, so we also discuss the factors that determine how widespread the representations of different objects will be, and whether binding might prove valuable by selecting subsets of active neurons for further processing. Finally, we consider whether binding by temporal correlation is feasible, addressing in particular some problems related to synchronizing sensory responses that are widely scattered in cortex. We will focus primarily on the neurophysiological data, leaving aside questions about the anatomical pathways and circuits that...
might bring about synchronization. While these are important topics, existing anatomical observations do not strongly constrain hypotheses about binding.

**Is Binding Needed?**

It has been suggested that binding must occur because otherwise there are too few neurons to support all possible percepts (Engel et al., 1992b, 1997; Singer and Gray, 1995). This argument is generally expressed in terms of a combinatorial explosion. If the presence of a colored bar of light were represented by the activity of an individual neuron, many neurons would be needed to represent all the possible combinations of orientations and colors of bars that might be seen. If 100 neurons were needed to represent all possible colors and 100 neurons were needed to represent all possible orientations, then 10,000 neurons would be needed to allow individual neurons to represent each possible combination of colors and orientations. To represent brightness in addition to color and orientation, the number would grow by another factor of 100, rising to 1,000,000. Because each stimulus dimension (e.g., length, width, etc., distance) increases the number of neurons exponentially, there could not be enough neurons to represent in this way all possible combinations of stimulus attributes.

Any stimulus combination might be represented, however, if all the neurons that were activated by one object could be bound together with some type of label. In that case 100 neurons signaling color, and 100 neurons signaling orientation, and 100 signaling each of the other relevant attributes, would suffice to completely specify any bar. Because the number of neurons rises arithmetically with the number of stimulus attributes, rather than exponentially, there is no combinatorial explosion if signals can be bound.

This line of reasoning suggests that binding of the sort provided by temporal correlation is needed, but is the combinatorial explosion a real problem? While our gaze might fall on an effectively infinite number of different objects, the critical question is not how many different objects might appear; rather, it is how many our visual system allows us to distinguish from one another. If two objects are not seen as distinct, there is no reason that they must have different representations in the nervous system. To the contrary, a failure to differentiate stimuli suggests that their neuronal representations do not differ in any important way.

The number of objects or items that humans distinguish is not known with any precision, but it is finite. It has been estimated that people distinguish fewer than 100,000 different types of objects (Biederman, 1987). The number of distinguishable items is obviously larger, because we recognize specific instances of some objects, and distinguish things that would not be counted objects, such as text, scenarios, and differences in such low-level attributes as orientation, brightness, and color. But even a generous factor of 100 would leave the final value for distinguishable visual items well below the number of neurons in the visual cerebral cortex. Consistent with this, psychophysical studies suggest that we do not distinguish a near-limitless number of objects. Untrained subjects do not see differences that are readily visible to trained observers (Goldstone, 1998). In fact, people are remarkably insensitive to differences between visual scenes or objects (Simons and Levin, 1997).

We can, of course, see differences between any pair of objects presented, even if they do not appear distinct at first glance. Drawings of two snowflakes may at first appear identical, but a careful examination of the details of their outlines may reveal a subtle difference. This discrimination, however, is a sequential process of evaluating the individual edges or angles, which themselves become the objects of scrutiny. A sequential process of this sort would require binding over time, but, as mentioned above, that form of binding is undisputed. Although we can detect many different attributes of an object, there is no reason to believe that all attributes need to be, or can be, detected simultaneously.

If the visual system needs to represent a finite set of objects and attributes, there could be small groups of specialized neurons for each discriminable object or attribute. Because the visual system may not need to access any of these different representations simultaneously, there may be no need for binding the activity of distributed neurons with synchronization.

**Specialized Neuronal Representations in Visual Cortex**

If binding is to be avoided, visual cortex must contain neurons with sophisticated receptive field properties matched to the objects and attributes that are readily discriminated. The suggestion that visual performance might depend on highly specialized neurons was formalized in the notion of “cardinal cells” proposed by Barlow (1972). The formulation most relevant to the issue of binding is the proposal (slightly modified from the original) that for an observer to distinguish two visual stimuli, there must exist in visual cortex a small group of neurons that do at least as well as the observer: that is, an analysis of their spike trains would yield performance equal to or better than the observer’s behavioral capability (Barlow, 1985). Although the term “cardinal cell” usually brings to mind neurons that respond to complex stimulus patterns or forms, we leave open the possibility that neurons throughout visual cortex could contribute equally, so that neurons in earlier stages could be involved in discrimination of low-level attributes such as precise positions or orientations, and those in later stages involved in the assessment of complex patterns and forms. Because representation of discriminable objects by small groups of highly specialized neurons would greatly reduce the motivation for binding by temporal correlation, we turn now to the question of whether the neuroanatomical and neurophysiological data are consistent with such an arrangement.

Comparisons of behavioral performance and the responses of individual neurons support the idea that decisions and actions could be based on the activity of small groups of neurons in visual cortex. These studies have found that the most sensitive neurons carry signals that match behavioral performance (reviewed by Barlow, 1972, 1985; Parker and Newsome, 1998). Investigations of this sort have addressed sensitivity to low-level attributes, such as discriminations of orientation, direction, or spatial offsets. While it remains to be seen
whether a one-to-one correspondence exists between behavioral and neuronal performance for visual discriminations of more complex stimuli, the existence of neurons in inferotemporal cortex that appear to reliably signal the differences between specific faces (Perrett et al., 1984) or other complex forms (Logothetis et al., 1995) suggests that this will be the case.

Could cortex contain enough small groups of neurons to account for the range of visual identification and discrimination that humans perform? The number of neurons appears sufficient. Visual cortex in both cerebral hemispheres of the human brain contains about one billion neurons (see Figure 1 legend), and we have argued that people do not discriminate or identify a billion objects or attributes.

Does visual cortex contain neurons with the sort of highly specialized response properties that would be needed for the range of discriminations that people can do? In macaque monkeys, the species for which we have the most detailed information, recordings from individual cells have shown that neurons in visual cortex have preferences spanning a broad spectrum. Areas at early stages of cortical processing contain neurons that encode simple visual attributes, and those at later stages contain neurons that encode more complex information about the visual scene (see Van Essen and Gallant, 1994). For neurons at intervening stages of visual cortex, our understanding of response properties is rudimentary, but they appear to provide representations of intermediate complexity (e.g., Kobatake and Tanaka, 1994; Gallant et al., 1996). Surveys of response properties across different cortical areas have described several gradual transitions going from earlier stages to later ones. These include gradual increases in receptive field size (see Maunsell and Newsome, 1987), the complexity of optimal stimuli (Kobatake and Tanaka, 1994), and the influence of extraretinal inputs in successive cortical levels (Motter, 1993; Ferrera et al., 1994; Treue and Maunsell, 1996; Luck et al., 1997a).

The organization of visual cortex is well suited to supporting specialized representations. It has long been recognized that visual cortex is subdivided into distinct areas, and that these can be grouped into functional streams that support different visual behaviors (reviewed by Merigan and Maunsell, 1993). Subdivisions exist on finer scales as well, so that cortex contains a hierarchy of representations with finer physiological differences on successively smaller spatial scales (Figure 1). Striking differences in stimulus preferences have been seen between the coarser subdivisions. The middle temporal visual area (MT) and area V4, although physically adjacent, lie in different streams. The earliest studies of these areas by Zeki and his colleagues (Dubner and Zeki, 1971; Van Essen and Zeki, 1978; Zeki, 1978, 1983) reported that neurons in MT and V4 could be sharply distinguished according to their selectivity for direction of motion (MT) and selectivity for color (V4). This strong distinction has since been confirmed in several quantitative studies (see Felleman and Van Essen, 1987).

Other observations have shown differences in response properties arising in the transition from one cortical area to another within a stream. One is the appearance of responses to complex pattern motion at the level of MT (Movshon et al., 1985; Rodman and Albright, 1989), which are not found in V1. Similarly, the transition from MT to the dorsal subdivision of the medial superior temporal area (MSTd) yields neurons that respond preferentially to patterns of optic flow (e.g., expansion or rotation) rather than simple translation (Saito et al., 1986; Tanaka and Saito, 1989; Tanaka et al., 1989).

Figure 1. Hierarchy of Spatial Scales of Organization in Visual Cortex

The approximate number of each type of subdivision is given for one cerebral hemisphere of a macaque monkey. These values are based on the following assumptions: there may be processing streams other than the dorsal and ventral streams (e.g., Boussaoud et al., 1990; the number of known visual areas (Felleman and Van Essen, 1991) will increase only moderately through subdivision of existing areas and assignments of unassigned regions; most modules will be as large or larger than the blobs in V1 (Tootell and Hamilton, 1989; Purves and LaMantia, 1990); columns are about 50 μm wide (Hubel and Wiesel, 1974); and about 10^7 neurons lie below each square millimeter of cortex surface, except in V1 where the number is about twice that (see Rockland, 1997). The level labeled "subcolumns" refers to a level of organization between that of a column and individual neurons. It corresponds to laminae or sublaminae within a single column. Each level is made up exclusively of elements from the level below, with the approximate number of contributing elements given between each pair of levels. These values are given as ranges in part because accurate numbers are not known (e.g., the exact number of areas in each stream of processing), and in part because the values vary within cortex (e.g., there are more columns in the cytochrome oxidase stripes that make up modules in V2 than there are in the cytochrome oxidase blobs in V1). As an approximation, the number of elements in one cerebral hemisphere of human visual cortex can be taken as twice as large from the level or modules or columns downward (Rockel et al., 1980; Tootell and Taylor, 1995; Tootell et al., 1996).
While these observations are consistent with the notion that each small group of neurons in visual cortex has highly specialized stimulus preferences, we must acknowledge that the supporting data are limited, and that many studies have found little difference between the response properties of neurons in different subdivisions of cortex. To some extent, this can be attributed to a paucity of relevant experiments. For example, only a handful of studies have compared directly the response properties of neurons in the different streams of processing. But most of these studies have failed to show striking differences. For example, shape selectivity, a property normally associated with the ventral stream, can also be found in areas of the dorsal stream (Sereno and Maunsell, 1998). Similarly, signals related to remembered direction of motion, a property normally associated with the dorsal stream, have been found in the ventral stream (Ferrera et al., 1994). Many anatomical pathways have been found to link the dorsal and ventral streams (see Merigan and Maunsell, 1993), raising the possibility that there might be considerable intermixing of their signals.

Clear specialization has been particularly difficult to demonstrate between finer spatial subdivisions, such as the modules that exist within areas. Although early reports emphasized a segregation of color and orientation selectivities between the blobs and interblobs in V1 (Livingstone and Hubel, 1984), subsequent measurements have not found a clear separation of these properties (Leventhal et al., 1995). Instead, there are general trends, such as neurons in the blobs having somewhat greater contrast sensitivity, and those outside preferring higher spatial frequencies (Silverman et al., 1989; Edwards et al., 1995). Similarly, when properties such as color, orientation, and direction selectivity are tested in the different V2 cytochrome oxidase stripes, trends are evident, but cells with a given particular preference can be found in any type of stripe (DeYoe and Van Essen, 1985; Peterhans and von der Heydt, 1993; Levitt et al., 1994; Roe and Ts'o, 1995).

While the results from many quantitative comparisons suggest that the similarities in the stimulus preferences of different groups of neurons generally exceed the differences, there are good reasons to believe that once response properties are fully understood, each neuron, or small group of neurons, will be found to have a unique stimulus preference. Perhaps the most important reason for this is the potential for overlooking or underestimating differences. The response properties of cortical neurons are proving to be complex and subtle, and there is no formula for finding an ideal stimulus. Finding stimuli that reveal differences in response properties is challenging. Until about 15 years ago, neurophysiologists who wanted to see a response property that did not exist in V1 had to go all the way to the final stages of visual cortex, where it was possible to find selectivity for complex patterns, such as faces, in inferotemporal cortex (Gross et al., 1972), or complex patterns of motion in parietal cortex (Motter and Mountcastle, 1981). Our understanding of this subject is now sufficiently refined that there are examples of new properties emerging from one visual area to the next, but still only a few.

Differences between neurons' response properties will be underestimated if they are not tested with the stimuli for which they have the greatest sensitivity. This is demonstrated clearly in comparing neurons in MT and V4. Although they are distinct in their selectivities for color and direction, quantitative comparisons of selectivities for stimulus orientation, speed, length, or contrast show little difference between these cells (Felleman and Van Essen, 1987; Cheng et al., 1994). If any pair of neurons is tested for sensitivity to inappropriate stimulus dimensions, their preferences will appear less distinct than they actually are. Because differences between nearby cells are most likely to be subtle, the problem is most acute at finer spatial scales, and few signs of specialization have been found within modules or columns. It is nevertheless likely that appropriate measurements will reveal specialization on even these scales. For example, when the receptive field structure of nearby V1 neurons is measured with high precision, there is generally little overlap (DeAngelis et al., 1999). Although the processing of signals within cortical columns remains poorly understood, there is little doubt that signals are substantially transformed as they pass between layers or sublayers within a column. It is ironic that columns are defined by a uniformity of response properties, yet most neuroscientists accept that the transformation of signals as they flow within a column may be one of the most important functions of cortex.

A further complication for recognizing neuronal specialization arises from the likelihood that differences between the response properties of many neurons may include extraretinal factors. Recent recordings from behaving animals have shown that neurons in different parietal areas can be distinguished by their activity in different behavioral contexts more clearly than they have been distinguished by stimulus preferences (Snyder et al., 1997; Gottlieb et al., 1998; Eskandar and Assad, 1999). Identifying appropriate behavioral contexts is at least as difficult as identifying appropriate retinal stimuli.

While the available data are far from conclusive, they support the suggestion that each small group of neurons in visual cortex has a unique stimulus preference. Thus, visual cortex may contain a specialized set of cells for each of tens of millions of different visual items, enough to match the number of different visual objects, forms, and attributes that people distinguish. If so, there may be no need for the type of binding that would be provided by temporal correlation.

An obvious question is how neurons in visual cortex could acquire the sophisticated receptive field properties needed to represent the specific types of visual objects that people discriminate. While a fixed set of response properties might serve for all low-level representations, there is no way to anticipate what complex patterns an observer might need to discriminate. In this regard, it is important to note that there is evidence that the stimulus preferences of cortical neurons are modifiable. Neurophysiological studies suggest that neurons change their stimulus preferences to match the properties of novel objects or features as subjects learn new discriminations, including the elaboration of complex response selectivities (Sakai and Miyashita, 1991, 1994; Logothetis and Pauls, 1995; Gibson and Maunsell, 1997; Kobatake et al., 1998). Plasticity in adult visual...
cortex could allow visual experience to shape the preferences of neurons to create representations that would support whatever discriminations are needed.

Would Binding Help?
Even if there is no absolute need for binding by temporal correlation, it might play an important role in visual processing. If a huge number of neurons are activated by a visual scene, then binding by temporal correlation might help by labeling a subset of neurons that the nervous system needs to evaluate. We consider now the question of how large a population of cortical neurons is activated by a visual object.

Even if each neuron has a unique stimulus preference, an object may activate many cortical neurons. There are two reasons why this is so. The first is that any object has many recognizable attributes, from the location, color, and orientation of each of its component edges and parts, to its overall pattern and form. Neurons that respond selectively to each of these attributes would all be activated. The second is that although the optimal stimulus for each neuron may be different, their selectivity for each stimulus dimension or attribute is broad. For example, individual orientation-selective neurons typically respond well to about 30°-50° out of a possible 180° (DeValois et al., 1982b). Similarly broad tuning exists for other stimulus dimensions, such as spatial frequency (DeValois et al., 1982a; Foster et al., 1985), binocular disparity (Poggio and Fischer, 1977), color (Lennie et al., 1990), and speed (Orban et al., 1986).

If neurons are broadly tuned for a given stimulus dimension, then a particular object might activate many neurons. Consider a population of neurons that was sensitive only to orientation, with each neuron preferring a different orientation. If these neurons are tuned so that each responds strongly over 30° (one-sixth of the range of possible orientations), then a stimulus of any orientation would correspondingly be expected to produce strong responses in one-sixth of the population (all neurons whose preferred orientation was within 15° of the orientation presented).

On the other hand, broad tuning for individual stimulus dimensions does not necessarily mean that many neurons will be activated by a single object. Sensory neurons are typically sensitive to several stimulus dimensions. Although the orientation of a given stimulus might be appropriate for many neurons, every stimulus has other attributes, such as color, spatial frequency, or binocular disparity. If neurons are sensitive to more than one stimulus attribute, they will respond strongly only to a small subset of possible stimuli, even if they are broadly tuned to each stimulus dimension. (DeAngelis et al., 1990). Imagine that the population of orientation-tuned cells considered above was also sensitive to spatial frequency, with each cell preferring a different spatial frequency and being as broadly tuned for spatial frequency as it is for orientation. Any given stimulus, now defined by both orientation and spatial frequency, will no longer activate one-sixth of the population but rather one-thirty-sixth of the population (1/6 × 1/6). As more and more sensitivities are considered, a smaller and smaller subset of the population will find a particular stimulus adequate. If each neuron were sensitive to about ten stimulus dimensions, then only a few neurons in all of visual cortex would be expected to respond to a particular combination of stimulus attributes.

The number of neurons that will be activated strongly by a stimulus therefore depends on both the breadth of tuning for stimulus dimensions and the number of stimulus dimensions to which each neuron is sensitive. Although definitive data on this issue would be difficult to obtain, we can infer the number of stimulus dimensions for which a neuron is likely to have substantial tuning. The most important observation is that it is possible to find an adequate stimulus for a sizable fraction of the neurons that are encountered in visual cortex. There has been no systematic study of the proportion of neurons for which an adequate stimulus can be found, but anecdotal observations suggest that most cells are not overwhelmingly demanding. A survey that tested every neuron isolated in V1 found that about 90% of the cells responded to a stimulus that was centered on their receptive field and adjusted only in orientation (Maunsell and Gibson, 1992). Even in inferotemporal cortex, where neurons can show remarkable specificity in their stimulus preferences (see Desimone, 1991), more than 70% of cells have been found to respond to stimuli drawn from a limited set (Desimone et al., 1984). The fact that adequate stimuli can be found for most cortical neurons suggests that, individually, they are not strongly selective for a large number of stimulus dimensions.

A combination of broad tuning and selectivity for a limited number of stimulus dimensions allows a large population of cortical neurons to be activated by any given stimulus, even when each has a distinct preferred stimulus. The size of the active population will depend on the particular stimulus. An object that covers all of the central visual field would probably activate millions of cells, owing to the magnification of this part of the visual field in cortex. It is unlikely that any stimulus would activate fewer than tens of thousands of neurons, or would activate neurons in only one visual area. The widespread patterns of activation that are often seen in experiments that image activity in human brain (Wandell, 1999) are consistent with this, although human imaging methods do not give precise information on the distribution of activity owing to limitations in sensitivity and vagaries in the relationship between regional blood flow and the distribution of active neurons.

Given that a visual scene will activate large populations of neurons in visual cortex, might temporal correlation help by labeling neuronal signals associated with one object in the scene, thereby limiting the number of cells that need to be analyzed? Temporal correlation would reduce the number of relevant signals, but it not clear that this would be especially valuable. Even if the labeled cells were one-tenth or one-hundredth of all active neurons, they will still comprise a large, widely distributed population. As mentioned above, the population could include neurons that were selective for every distinct attribute and aspect of the object, from the position and movement of every resolvable edge to its overall shape. Their activity could not simply be summed or averaged. Some process would still have to extract the meaningful signals from the pattern of activity. How any distributed patterns of activity, whether they are among hundreds or millions of neurons, can lead to unitary
percepts or singular actions remains a mystery. Binding the activity evoked by particular objects does not solve this problem and does little to reduce its scope. The gain from binding with temporal correlation would be quantitative, not qualitative.

**Difficulties Introduced by Temporal Correlation**

Because temporal correlation requires some temporal precision in the firing of neurons, it is important to consider the feasibility of creating and detecting correlated firing among widely scattered neurons. Response oscillations at moderate frequencies will interfere with time-varying sensory inputs, and synchronizing the sensory responses of extended populations of neurons may be difficult to achieve.

Most studies of the synchronization of neuronal responses have used stimuli that are either stationary or moving at a constant speed, but neurons can represent more dynamic stimuli with high temporal precision. When neocortical neurons are repeatedly stimulated with a varying input in vitro, they can produce spike trains in which the timing of spikes is consistent between repetitions to a precision of a few milliseconds (Mainen and Sejnowski, 1995). Studies in vivo have shown that neurons in MT similarly can convey signals with timing that is reliable to a precision of a few milliseconds between stimulus repetitions (Bair and Koch, 1996; Buračas et al., 1998). Psychophysical observations suggest that the temporal precision of cortical neurons is important for guiding behaviors. Humans are sensitive to interocular timing differences as small as 150 μs (see Barlow, 1981). Because signals from the two eyes are first combined in visual cortex, some cortical neurons must convey signals of extraordinary precision.

The existence of excellent temporal precision in visual cortex poses a problem for binding groups of neurons using temporal synchronization. If cells encode moment-to-moment changes in a stimulus with a precision of a few milliseconds, information will be lost if their signals are modulated at frequencies below 100 Hz, as proposed by the temporal correlation hypothesis. The timing of spikes cannot easily be used to encode information about binding when spike timing is also needed to encode information about temporal variations of visual stimuli. Although such high temporal precision may be limited to certain types of stimuli, its existence implies that there are situations where binding based on temporal modulations of responses would compromise visual signals.

Other problems for synchronization may arise from the surprisingly long time that it takes for visual signals to propagate through cortex. Average latencies for responses in V1 are usually reported in the range of 50–80 ms (Maunsell and Gibson, 1992; Nowak et al., 1995a). Average latencies in V2 are about 5–15 ms longer (Rai-guel et al., 1989; Nowak et al., 1995a; Schmolesky et al., 1998), and those in inferotemporal cortex are about 40 ms later that those in V1 (Baylis et al., 1987; Vogels and Orban, 1994; see also Schroeder et al., 1998). Slow propagation is also seen in the relay of signals between layers within individual areas. In V1, the most superficial neurons do not become active until about 10–15 ms after neurons in layer 4 (Best et al., 1986; Maunsell and Gibson, 1992; Schroeder et al., 1998). Differences in response latencies are likely to be even larger when viewing natural scenes. A neuron’s latency varies depending on the stimulus presented, with weaker stimuli typically yielding longer latencies and larger differences between neurons (Levick and Zacks, 1970; Shapley and Victor, 1978; Lennie, 1981; Sestokas and Lehmkuhle, 1988; Givre et al., 1995; Maunsell et al., 1999; but see also Gawne et al., 1996). Because stimuli that are optimal for neurons in the early stages of visual cortex may not be optimal for neurons in later stages, in some situations the time it takes for detectable signals to propagate across visual cortex may approach 100 ms.

The large timing differences that exist across visual cortex are not a major issue for conventional views of cortical processing, in which each neuron samples inputs from cells with related response properties that are generally dispersed across only a few cortical levels (Marsálek et al., 1997). But large timing differences are likely to impose limitations if binding by response modulation is needed for visual processing. Activation of cells throughout visual cortex, including the slowest to respond, may be required before sufficient information to support appropriate binding is available. The loss of early, unsynchronized responses could be a significant penalty given the small number of spikes that cortical neurons produce during behaviorally relevant intervals. Behavioral decisions are typically made within a few hundred milliseconds, during which even a strong response would amount to no more than 10 or 20 spikes per cell.

Timing differences may also severely limit the duration of synchrony both within and between areas. Cells with early, transient responses may stop responding before others begin to respond strongly (Schmolesky et al., 1998). For briefly presented stimuli, the responses of all V1 neurons may be nearly complete before the final stages of visual cortex have well-established responses (Givre et al., 1995). Neurophysiological measurements of synchronization have typically used bright, uniformly illuminated stimuli on uncluttered and featureless backgrounds. The results may not generalize to more natural viewing conditions, in which neuronal responses to different parts of a particular object might overlap little in time. For example, parts of an object that are in shadow would be expected to generate cortical responses that lag far behind those produced by parts in bright light.

**Concluding Comments**

Although we can infer how information about the visual scene is likely to be distributed in the activity of neurons across visual cortex, we still lack a sound mechanistic framework for explaining how these patterns of activity are converted into perceptions and behaviors. The temporal correlation hypothesis proposes a specific mechanism that might help in this process. It is not clear, however, that the problem that temporal correlation is supposed to solve, a numerical mismatch between the number of possible percepts and the number of cortical neurons, actually exists.

We have addressed some complications associated with binding by temporal correlation. These include the loss of signals about stimulus variations on fine time
scales and potential difficulties in dealing with substantial differences in response timing across large distances in cortex. There may also be a price to be paid in added neuronal machinery to identify neuronal signals that need to be grouped and to synchronize arbitrary pairs of neurons. According to the temporal correlation hypothesis, the purpose of these unspecified neural mechanisms is to label a subset of neurons as relevant. But even if the labeled cells are a fraction of all active neurons, they will still make up a large, widely scattered population. Temporal correlation would not solve the problem of unitary perception; it would only change the number of neurons that need to be processed. This seems a limited benefit, given the costs in terms of addition machinery and new complications to be addressed.

Whether temporal synchronization, or any other process, serves to bind signals distributed across cortex may not be resolved conclusively until we understand how distributed sensory representations are converted into actions. While that understanding may be distant, experiments show that the performance of individual neurons can equal behavioral performance, and that the cortex contains highly specialized representations that appear well matched to the range of visual discriminations that must be performed. It is therefore possible that much of visual performance relies on signals from relatively small groups of neurons, rather than widely distributed groups.

Acknowledgments

The authors thank C. E. Boudreau, E. P. Cook, M. C. Crair, J. J. DiCarlo, R. A. Eatock, and D. L. Sparks for thoughtful comments on the manuscript. Research in the authors’ laboratory is supported by grants from the Human Frontier Science Program and the National Institutes of Health (R01 EY05911). J. H. R. M. is an Investigator with the Howard Hughes Medical Institute.

References

A comprehensive reference list for all reviews can be found on pages 111-125.
References for Reviews on the Binding Problem


