The Temporal Correlation Hypothesis of Visual Feature Integration: Still Alive and Well

Review

Charles M. Gray*
The Center for Neuroscience and
Section of Neurobiology, Physiology, and Behavior
University of California, Davis
Davis, California 95616

In recent years, the binding problem has gained widespread attention among theorists and experimentalists alike. This surge of interest has been driven by the realization that there are few, if any, places in the nervous system where all the information necessary to carry out a particular task is localized. The inescapable conclusion is that sensory, cognitive, and motor processes result from parallel interactions among large populations of neurons distributed among multiple cortical and subcortical structures. Binding, in a general sense, is the process responsible for functionally linking this distributed activity. In the visual system, this problem corresponds to identifying the relationships among features in an image so that objects can be recognized. Much of this process relies on focused attention. It has also been recognized, however, that an early preattentive mechanism operates rapidly, and in parallel, to segment visual scenes into their candidate objects. This mechanism, often referred to as perceptual grouping, enables features to be grouped together according to Gestalt, and other, criteria. These processes are also faced with a binding problem. Some form of representation must be available to establish relationships among features and separate one set of grouped features from another. Milner (1974) and von der Malsburg (1981, 1985) proposed that these problems could be overcome if constellations of grouped features were represented dynamically by the formation of cell assemblies, defined by the selective synchronization of distributed neuronal activities. Such a mechanism, they argued, would permit the segregation of one representation from another and provide highly salient neuronal signals to an object-based attentional selection mechanism.

In the ensuing years, evidence has accumulated to support this hypothesis, but critics have argued that it is insufficient to warrant the theory's acceptance. They have also argued that the theory is flawed on physiological grounds, that cortical networks do not possess the requisite mechanisms to generate and make use of selective patterns of neuronal synchronization. In this review, I will argue to the contrary that the theory, particularly as it applies to preattentive vision, is well grounded in the psychophysical and neurophysiological domains. I will focus on the problem of visual perceptual grouping and define the process within the context of a set of principles articulated by Gestalt psychology. I will lay out the framework of the theory for how perceptual grouping is implemented in the visual system. And I will present arguments, couched in neurophysiological

terms, for why I believe this theory is attractive and has gained widespread attention. I will follow with a brief review of the evidence, both for and against, the theory and discuss some of the objections commonly put forth by its critics, in an effort to account for why the theory has been so attractive and so controversial at the same time. I will end with a discussion of specific experiments that are needed to rigorously test the theory and briefly summarize where it stands at present.

Binding Is a Diverse Set of Functions

Binding isn't a problem for nervous systems, as evolution has sculpted their organization to solve the problem efficiently and effectively. It is only a problem for those of us trying to understand how the nervous system achieves the task. Part of our difficulty stems from the fact that binding is a broadly defined term. It is a function that occurs over multiple spatial and temporal scales of organization in the nervous system. It can be rapid and preattentive or may require focused attention. It can occur within a single sensory modality or may require cross-modal or sensory–motor integration. Consequently, confusion and disagreement can easily arise among competing theories of the binding process and experimental tests of those theories.

To illustrate the richness of the binding process, consider a simple everyday example of a sensory–motor task. You want to find a certain object that's located on one of several adjacent shelves, but you're not sure which shelf, and there are many other objects on the shelves besides the one you're searching for. To find it, you have to visually scan the scene until you recognize the object. At each location of gaze, your visual system must parse the scene into its component objects. This involves identifying which features in the scene belong to which objects. Once this is accomplished, or perhaps at the same time, the identity of a particular object can be evaluated. If it doesn't match what you're searching for, you must direct your gaze to a new location and repeat the process. This redirection of gaze may itself be guided by the prior parsing of the scene. Once you've identified the correct object, you can reach out, grasp it, and carry on with the task at hand.

It is not entirely clear which components of this task we should refer to as binding and hence which neuronal mechanisms solve the "binding problem." Perhaps the most common view is that binding is the process of combining stimulus features to form an object representation. This is intuitively appealing, because it is well established that the locations and properties of local stimulus features are represented within multiple retinotopic cortical maps. The features comprising objects must somehow be "bound" together by combining the appropriate activities within the relevant cortical maps. However, a number of other processes in the task described above can be thought of equally well as binding. Associating the location of an object in space with its identity is one example. Abundant evidence indicates that separate cortical areas participate in the recognition
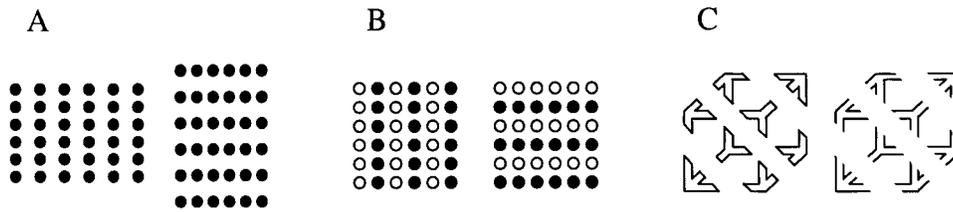
*E-mail: cmgray@ucdavis.edu.

Figure 1. Simple Patterns Illustrating Several Gestalt Principles of Perceptual Grouping

(A) Proximity. The arrays of dots are automatically grouped into columns or rows depending on the nearest neighbor relationships.

(B) Similarity. Arrays of equally spaced dots are automatically grouped to form columns or rows depending on the axis of similarity of the features.

(C) Closure and collinearity. The diagram on the left appears as a collection of unrelated objects. If a subset of the lines are removed, the collinear alignment of the remaining lines reveals a 3-dimensional Neckar cube and several overlapping illusory contours. Adapted from Kanizsa (1979).

and spatial localization of objects. Hence, the need arises to bind the neuronal activities representing "where" an object is with those representing "what" it is. Another compelling example is the association between the sensory representation of objects and the decisions and motor commands to act in relation to those objects. These are functions that clearly involve a coordinated interaction between distributed regions of the brain involved in sensory, cognitive, and motor processing. Thus, it seems reasonable to think of many neuronal functions involving the integration of distributed information as binding processes.

Such a general definition, however, might lead to confusion between processes mediated by different neuronal systems. For example, some binding processes, such as the grouping and segmentation of features, take place preattentively, while others, such as object recognition and aspects of sensory–motor integration, require focused attention. To lump these processes together under the term "binding," and to then set out to solve the "binding problem," would appear misguided. Because these processes are so diverse, it is likely that they are mediated by a variety of mechanisms. As a result, we may introduce difficulties (and controversy) in understanding binding mechanisms if we do not constrain the problem.

Perceptual Grouping Is a Form of Binding

Every time we inspect a visual scene, we are confronted with a complex pattern of features, some static and some moving in various directions and at various velocities. In spite of this complexity, our visual system is able to rapidly identify the features that belong together as objects. In the first half of this century, the field of Gestalt psychology formulated a set of principles specifying the properties of visual features that observers tend to group together (Wertheimer, 1923, 1955; Köhler, 1930; Koffka, 1935, 1969; Kanizsa, 1976). These were features of high visual salience that included their so-called laws or "factors" of proximity, similarity, continuity, closure, and common fate. They discovered that features with these common characteristics tend to be automatically "bound" together as coherent units (Figure 1) and thus often define objects in the real world. This made intuitive sense, given that visual objects tend to be composed of features that are spatially adjacent, have similar appearance, are continuous with one another, form closed

contours, and move coherently together. It should be noted that the characteristics listed above are not the only stimulus features that contribute to grouping. Numerous studies have demonstrated that a variety of other stimulus features, such as size (Bergen and Adelson, 1988), texture (Julesz, 1975), binocular disparity (Nakayama and Silverman, 1986), and coincidence in time (Alais et al., 1998) also contribute to grouping.

The Gestalt psychologists postulated that sensory grouping is an automatic and dynamic process that binds sensory units (features) together into emergent wholes or objects. Grouping was considered to be an early step in perception that preceded recognition. Support for these arguments gained ground when it was discovered that under certain conditions the time required to identify a target in a display did not depend on the number of items in the display (Neisser et al., 1963). Results of this type led to the notion that early visual processing does not depend on a serial process mediated by directed attention. This "preattentive processing" was postulated to occur in parallel over the whole visual field and to provide an initial segmentation of a visual scene into its component objects on the basis of Gestalt properties (Neisser, 1967). Once completed, the signals resulting from the grouping process could then serve as inputs for further analysis using selective attention. Such a mechanism would dramatically reduce the number of elements the attentional system would be required to scrutinize and thereby increase the speed and efficiency of visual search.

In the years since, numerous studies have provided support for the general framework of this hypothesis (Rock and Brosgole, 1964; Beck, 1966, 1967; Kahneman, 1973; Julesz, 1975; Kahneman and Henik, 1981; Pomerantz, 1981; Prinzmetal, 1981; Treisman et al., 1982; Bergen and Julesz, 1983; Duncan, 1984; Bergen and Adelson, 1988; Bravo and Blake, 1990; Elder and Zucker, 1993). In what have come to be known as object-based models, attention is thought to be directed to collections of features that are formed preattentively into primitive objects (Neisser, 1967; Duncan, 1984; Donnelly et al., 1991; Vecera and Farah, 1994; Rensink and Enns, 1995; Wolfe and Bennett, 1996). This suggests that perceptual grouping is an early form of feature binding that identifies and represents relationships among features, and that these representations can be accessed by attentional selection mechanisms. This notion leads to a number of obvious questions. What are the features that are
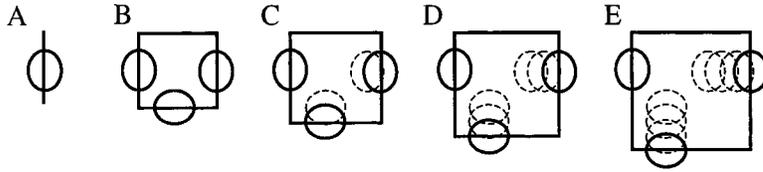
Figure 2. Schematic Diagram Illustrating the Requirement for Flexibility in the Grouping of Visual Features

(A) A vertically oriented contour centered over the receptive field of a hypothetical group of visual cortical neurons.

(B) Grouping of this contour with other local contours to form a square requires the integration of signals from other cortical neurons that are activated by the other parts of the figure. Two of the hypothetical receptive fields are included.

(C–E) The remaining plots illustrate that there exist a large number of other possible squares that contain a vertical contour overlying the receptive field shown in (A). In each of these examples, the cells activated by this contour must now integrate their activity with those activated by different sets of contours whose receptive fields lie in distinct positions. The dashed ellipses in (C) through (E) illustrate the cumulative positions of the receptive fields for each of the objects.

preferentially grouped together? What type of representation is formed among the grouped features? What are the neurophysiological constraints within which these processes must operate? And what aspects of the representation are prerequisites for attentional selection?

I will argue, as have many others, that these primitive, preattentive objects result from the grouping of features according to Gestalt principles. For simplicity, I will limit my arguments to those feature elements that contribute to form and shape perception. I will further argue that the representation of grouped features must rely in some way on the formation of neuronal assemblies (Hebb, 1949; Braitenburg, 1978; Edelman, 1978; Palm, 1981, 1990; von der Malsburg, 1986; Gerstein et al., 1989). These dynamic structures provide the most parsimonious mechanism to account for the rapid, parallel, and combinatorial nature of perceptual grouping. They also provide a neurophysiologically plausible mechanism for greatly increasing the salience of neuronal activity for subsequent processing stages involving selective attention (Reynolds and Desimone, 1999 [this issue of *Neuron*]). To fully explore these arguments, it is necessary to first describe the properties of perceptual grouping so that the neurophysiological constraints on the task are fully appreciated.

Properties of Perceptual Grouping

Although perceptual grouping appears to occur effortlessly, several aspects of the process illustrate why it is so difficult to understand mechanistically, and why "binding" has become such a "problem." First and foremost, perceptual grouping is a flexible process. A given feature element, at a given location in the visual field, may be grouped with one set of elements in one scene and with another, completely different, set of elements in another scene. To illustrate this point, consider the simple diagrams shown in Figure 2. A single vertically oriented contour, located at a fixed position in space (Figure 2A), can be combined with other contours to yield the shape of a square (Figure 2B). If the size of the square changes (Figures 2C–2E), while the position of its upper left vertical contour does not change, then the same vertical contour must be combined with a different set of contours for each configuration of the square. This implies that the neuronal activities signaling the presence and properties of the vertical contour must be combined with different populations of active neurons as the properties of the image change. Taking this simple argument for flexibility one step further, it is easy to

appreciate that the combinatorial capacity of perceptual grouping must be very high. If the vertical contour in Figure 2A is combined with contours of different orientations and positions, a wide variety of shapes are possible (Figure 3). As the relationships among the features change, the grouping process must adapt to cope with the large number of possible combinations. Thus, the activity signaling the vertical contour must be recombined in a flexible manner so that a large number of other feature relationships can be perceptually grouped. Given that visual scenes exist in near infinite variety and constantly appear in novel configurations, the perceptual grouping process appears, for all practical purposes, to have an infinite capacity.

Another readily apparent feature of perceptual grouping is its parallel organization. Features comprising figures, such as those shown in Figures 2 and 3, rarely exist in isolation. Rather, visual scenes are usually composed of complex collections of objects that often overlap or occlude one another. This implies that the grouping of features must take place simultaneously at multiple, and sometimes overlapping, locations of the visual field. When this requirement is combined with the need for flexibility and combinatorial capacity, it is easy to appreciate the inherent complexity of the neuronal processes that must be involved. Multiple sets of features must be grouped simultaneously and perceptually segmented from one another. However, as the visual scene changes, features that are grouped together in one image may be segregated from one another in a different image. This may occur even if the properties of the local features themselves do not change. This notion is illustrated in Figure 4. The plots in Figure 4A show the receptive fields of six hypothetical cortical neurons, each tuned to vertical orientation. One can imagine many constellations of features in which all six of the cells are activated by their preferred local features. The plots in Figures 4B–4D illustrate three such examples. What distinguishes these simple scenes from one another are the obvious differences in perceptual grouping that contribute to the recognition of either two or three distinct figures. In each image, a different pattern of grouping must take place while the local features present in the receptive fields of the cells remain roughly constant. For example, two vertical contours are grouped when they form opposite sides of a square but are segregated from one another when they form components of two different squares. These examples illustrate how perceptual grouping must be both flexible and parallel in
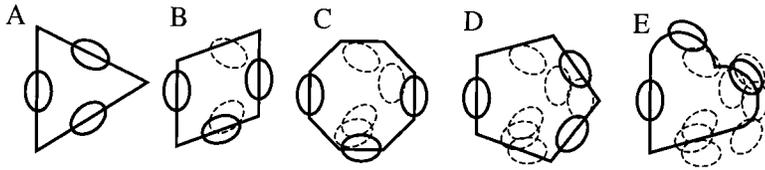
Figure 3. Another Example Illustrating the Combinatorial Complexity of the Grouping Process

The activity of the cells signaling the vertical contour on the left of each figure must be combined with the activity of cells signaling a large number of other possible features. For any given object, a distinct grouping operation should take place. This indicates that the visual system must be capable of identifying the relationships among a large number of possible feature combinations. For each of the figures, a set of three hypothetical receptive fields are shown. The dashed ellipses in (B) through (E) illustrate the cumulative positions of the receptive fields for each of the objects.

its organization, but they also suggest that competition may occur among grouping criteria such that certain feature relations, like continuity and closure, may tend to dominate over others (Donnelly et al., 1991; Elder and Zucker, 1993; Kovacs and Julesz, 1993).

Finally, another aspect of perceptual grouping, its speed, reveals the truly impressive nature of the process. It is not certain how long a given grouping operation takes to complete, and the duration no doubt varies with the properties of a scene, but estimates usually fall in the range of 100–300 ms (Beiderman, 1981; Intraub, 1981; Thorpe et al., 1996). This appears reasonable on several grounds. When observers scan visual scenes they usually make saccadic eye movements anywhere from one to four times a second, with the duration of visual fixation occasionally being as brief as 100 ms. Because the retinal image changes with each new visual fixation, perceptual grouping should take place at least as quickly as the interval between saccadic eye movements. Evidence for this comes from experiments in which observers, signaling the presence or absence of an animal in briefly presented scenes, exhibit event-related potential components that correlate with task performance as early as 150 ms after stimulus onset (Thorpe et al., 1996).

Together, these considerations suggest that the neuronal mechanisms responsible for perceptual grouping must operate within specific constraints. The mechanism must be capable of evaluating the relations among visual features and of identifying those combinations that exhibit Gestalt properties. It should be flexible, have a large combinatorial capacity, and be parallel in its organization so that many feature combinations can be

grouped simultaneously. The process must be fast, and should occur within the time available during the course of a visual fixation.

Potential Mechanisms Underlying Perceptual Grouping

What types of mechanisms might mediate the neuronal operations underlying perceptual grouping? Fortunately, research on the physiology and anatomy of the visual system has been one of the most active areas of neuroscience, so there are several good candidate mechanisms to consider. Here, I will briefly discuss three possibilities and consider their merits as well as their disadvantages.

Convergent Hierarchical Coding

The most widely accepted model for feature integration relies on the well-established understanding that visual cortex is organized hierarchically into a collection of distinct areas (Rosenquist, 1985; DeYoe and Van Essen, 1988; Felleman and van Essen, 1991; Sereno and Allman, 1991; Payne, 1993). In its simplest form, the visual cortical hierarchy can be viewed as a series of processing stages, where each stage carries out specialized, parallel operations that depend on the input from earlier stages. This hierarchical processing leads to an increase in the complexity of neuronal representations as the hierarchy is ascended. Evidence for this is revealed by the gradual decrease of retinotopic specificity, the increase in receptive field size, and the dependence of neuronal responses on increasingly complex stimulus features. At the same time, there is substantial evidence that separate cortical areas process specific attributes of visual features, suggesting that the analysis of visual
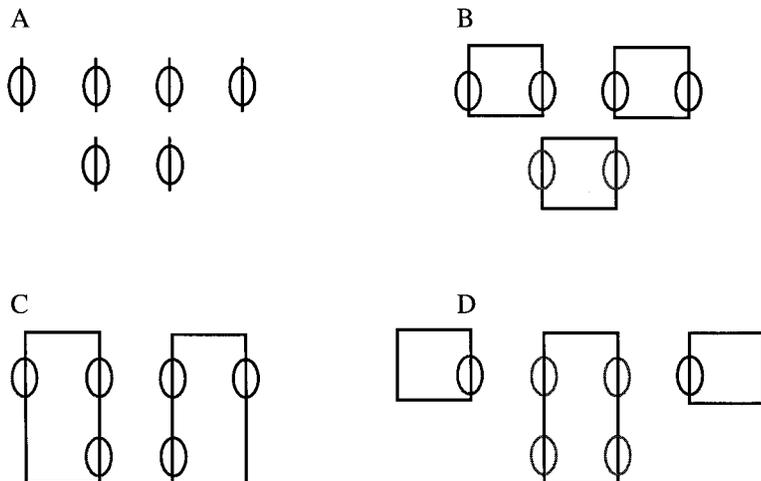

Figure 4. Diagram Illustrating the Requirement for Parallel Grouping and Segmentation

(A) Six hypothetical receptive fields, each tuned to vertical orientation.
(B–D) Three different scenarios in which the grouping criteria change, but the local features stimulating the cells remain relatively constant.

scenes is subdivided (Livingstone and Hubel, 1988; Zeki and Shipp, 1988; Van Essen and Anderson, 1990; Van Essen et al., 1992; Merigan and Maunsel, 1993). In primates, this organization culminates in the broad anatomical and physiological subdivisions of the dorsal and ventral streams, lying in the parietal and inferotemporal (IT) areas, respectively. These regions make fundamental contributions to the spatial localization and recognition of objects (Ungerleider and Mishkin, 1982; Goodale and Milner, 1992; Milner and Goodale, 1993).

This organization, particularly that of the temporal stream, has specific implications for preattentive processing. It suggests that grouping operations may be expressed by the activity of small collections of neurons (by single neurons in the extreme) by virtue of the elaboration of receptive field properties that occurs as the cortical hierarchy is ascended. This concept, most clearly articulated by Barlow (1972) and driven by the receptive field paradigm in visual neuroscience, implies that features and constellations of features are represented by the activity of neurons that receive convergent inputs from populations of neurons at lower levels in the hierarchy (Fujita et al., 1992; Tanaka, 1993). This type of mechanism is attractive for at least three reasons. First, there is a large body of evidence demonstrating that receptive field properties exhibit the requisite increases in complexity and specificity across the cortical hierarchy. At the highest levels, neurons exhibit selective responses to object-like stimuli (Gross et al., 1972; Desimone et al., 1984, 1985; Baylis et al., 1985; Perret, 1987; Fujita et al., 1992; Tanaka, 1993, 1996; Logothetis et al., 1995). Second, the selectivity for complex features is established early on in the time course of responses to stimuli (Rolls and Tovee, 1994). This indicates that the convergence of information, and hence the grouping of features, can satisfy the strict temporal requirements mentioned earlier. Finally, there are a number of studies demonstrating a close correlation between the activity of single neurons and the performance of animal subjects on sensory discrimination tasks (Newsome et al., 1989; Britten et al., 1992, 1996; Parker and Newsome, 1998). The combination of these attributes provides a strong argument that perceptual grouping can be achieved by hierarchical and convergent connectivity.

For a number of reasons, however, this is not likely to be the whole story. Convergent coding mechanisms of this type suffer from what is known as a "combinatorial problem." To appreciate this problem, it is useful to refer to the diagrams shown in Figure 2. In order to build a receptive field to code for the small square in Figure 2A, it would be necessary to send the outputs of the neurons coding for the orientations and positions of the local contours to a common location at a higher stage in the hierarchy. Neurons receiving these inputs could then represent a square at that location because their receptive fields would be organized to respond to the appropriate input. A similar mechanism could be used to represent the other squares in the figure. However, each of the squares includes a contour that lies over a common receptive field (upper left vertical contour). This means that the higher-level neurons, whose receptive fields code for the other squares, would all have to receive input from the neurons activated by the common

vertical contour. One can see that this type of mechanism quickly becomes untenable, when it is realized that the number of objects capable of sharing a common contour is very large. Each neuron would have to send an equally large number of connections to the next successive stage of the hierarchy, leading to an unacceptably large number of connections. Even if such connections could be configured appropriately, similar arguments apply to the number of neurons needed to represent all of the possible feature combinations. A mechanism of this type, in such a simple form, seems unlikely to solve the problem.

Population Coding

It has been recognized by many investigators that the combinatorial problem can be overcome by a simple modification of the convergent coding scheme outlined above. Rather than represent the integration of features by the activity of a few or even single neurons at specific cortical loci, complex feature combinations could be represented by the activity of a population of neurons distributed within and across levels of the cortical hierarchy (Ballard et al., 1983; Van Essen and Anderson, 1990). In this scheme, each stimulus pattern could be represented by a distinct pattern of firing in a distributed population of cells. Such a coding scheme would greatly increase the representational capacity of the cortical network, because the number of distinct patterns of activity far outstrips the number of neurons available to represent the stimuli. Thus, the combinatorial complexity of the sensory world would no longer pose a problem.

This scenario meshes well with many aspects of the anatomical and physiological organization of visual cortex. The anatomical connectivity within and between different levels of the cortical hierarchy is both divergent and convergent (Felleman and Van Essen, 1991; Rockland and Van Hoesen, 1994; Salin and Bullier, 1995; Levitt et al., 1996). Neurons in a given cortical area send divergent outputs to other areas, enabling the cells in these areas to receive convergent input from many sources. This type of organization provides neuronal populations at the highest levels of the hierarchy with broad access to the entire visual field and the full range of visual feature properties, clearly marking them as candidate areas for feature integration and perceptual grouping. Similarly, there is substantial physiological evidence from both sensory and nonsensory cortical areas that stimuli, as well as motor actions, are represented by distributed populations of cellular activity (Lee et al., 1988; Georgopolous et al., 1992; Young and Yamane, 1992; Wilson and McNaughton, 1993; Nicolelis and Chapin, 1994; Vaadia et al., 1995; Nicolelis et al., 1998). Thus, on both anatomical and physiological grounds, it appears reasonable that processes involved in sensory representation and perceptual grouping involve distributed neuronal activity.

Attractive though it may be, this model of sensory representation is again incomplete. To illustrate this, first consider a somewhat artificial situation in which a complex 3-dimensional object is presented against a background of uniform luminance and wavelength, a situation not unlike seeing a bird flying on a cloudless day. It is easy to appreciate how such a stimulus could give rise to a unique pattern of activity distributed across

the cortical hierarchy. The situation becomes considerably more complex when two objects are present in the same scene, particularly if they are close to or overlapping with one another. If each object gives rise to what would otherwise be a unique pattern of activity, what would identify the activity of the widely distributed neurons as belonging to one representation or another? Now consider the more common situation of a scene containing multiple objects and backgrounds varying in texture, luminance, and wavelength, and it's easy to see that a real conceptual problem arises. If features and objects are represented solely by the firing of distributed populations of cells, how are the vast numbers of active neurons in the cortical network disambiguated from one another? How can one or more distinct patterns be identified from the many others that are present in the same networks at the same time? This conundrum, first identified and articulated by Milner (1974) and von der Malsburg (1981, 1985), is often referred to as the "superposition problem."

The essence of this problem is one of identification and interference. How are members of a representation identified as belonging to that representation, and how is interference between simultaneous representations avoided? These questions lie at the heart of the binding problem and perceptual grouping. In models of attentional selection, these problems are often dealt with by saying that attention solves the binding and superposition problems by suppressing the activities evoked by all of the unattended features or enhancing the activities evoked by the attended features (e.g., Olshausen et al., 1993). However, this begs the question of how attention could be capable of serially searching all the possible feature combinations in a reasonable amount of time (Milner, 1974). Some form of image segmentation should occur that not only happens prior to attention but also serves to attract it (Neisser, 1967). One possibility is that salient features in an image, such as regions of high contrast or movement, lead to elevated activity, and the population of cells with the greatest activity captures attention (Koch and Ullman, 1985). The problem here is how to define the population. There may be multiple overlapping populations of cells with elevated activity that coexist in the same network. What kind of mechanism could serve to identify the appropriate features as belonging together?

Representation by Assemblies and the Temporal Correlation Hypothesis

A theoretical resolution of these problems was put forth independently by Milner (1974) and von der Malsburg (1981, 1985) (for related sets of ideas, see also Braitenberg, 1978; Edelman, 1978), which I will refer to as the "temporal correlation hypothesis" (Singer and Gray, 1995). The key insight in each of these models is simple. Given that the activity evoked by the features comprising an object is distributed, some mechanism is necessary to identify the members of a representation as belonging together and to distinguish them from other representations that may be present at the same time. Relying on an earlier set of ideas put forth by Hebb (1949), it was postulated that perceptual grouping and segmentation could be mediated by the formation of cell assemblies. This process was envisioned as occurring in parallel, such that multiple cell assemblies could be formed simultaneously, one for each set of grouped features. The properties of the cell assembly differed from those postulated by Hebb (1949) in that they were considered to be dynamic entities. The structure of an assembly would be determined not only by the anatomical organization of the cortical network, but it would also be constrained by the Gestalt properties of the stimulus features present in the scene. Because cell assemblies are defined as unitary entities, in which the neurons within an assembly interact more strongly with one another than with neurons outside the assembly (Braitenburg, 1978; Palm, 1981, 1990; Gerstein et al., 1989), this mechanism provided a means for the grouping of distributed activities and avoidance of the superposition problem. The question remained as to what mechanism would drive assembly formation, and what internal dynamic structure of a cell assembly would distinguish it from the other neuronal activity present simultaneously in the cortical network.

The common insight put forth to solve this problem was to define cell assemblies on the basis of the temporal correlation or synchronization of neuronal firing (Milner, 1974; von der Malsburg, 1981, 1985). Because neurons are sensitive to the timing of their synaptic inputs, it was argued that synchronization provided at least two important advantages. It would serve to reinforce the interactions among the members of an assembly and would provide a highly effective neuronal signal to systems responding to the output of an assembly. Although Milner (1974) did not explicitly define the time scale of the neuronal interactions within an assembly, von der Malsburg (1981, 1985) argued that synchronous firing should occur on a time scale of <5 ms. He also postulated a process of fast synaptic modulation, lasting hundreds of milliseconds, in which the strength of synaptic excitation onto a neuron would be transiently enhanced when the synaptic input to a neuron was synchronous with its spike output. This was postulated to provide a positive feedback mechanism to amplify weak correlations of activity. Milner (1974) recognized the need for a reinforcing mechanism and postulated that this function might be implemented through the feedback pathways from higher levels in the cortical hierarchy.

Given that perceptual grouping occurs over a range of spatial scales, both authors recognized the need for an anatomical substrate to mediate horizontal interactions within the cortical network. Milner (1974) postulated the existence of "distance cells" to link separate regions of the visual field, and von der Malsburg (1981) postulated a horizontal fiber system within primary visual cortex that would link populations of cells sensitive to the same categories of local features. In the latter scheme, common features in an image would lead to coactivation of cells with similar receptive field properties. Because of their lateral interconnections, these cells would in turn mutually reinforce one another. The resulting temporal correlations of firing would then be amplified by the fast synaptic modulation. In this way, assemblies could emerge through cooperative interactions, with their spatial structure being determined by both the stimulus as well as the underlying pattern of horizontal connections in the cortical network. (It is

worth noting that at the time these theories were developed very little evidence was available for the existence of horizontal connections in primary visual cortex [Creutzfeldt et al., 1977; Gilbert and Weisel, 1979], and there was essentially no evidence to suggest that the horizontal connections might link up populations of cells having similar receptive field properties. Both of these phenomena are now firmly established.)

Both authors proposed that this process should have hierarchical depth in the sense that neuronal assemblies would be defined not only within a given cortical area but also across levels in the cortical hierarchy, mediated through both feedforward and feedback pathways. Assembly formation would allow for local relationships to be defined early in the visual pathway, where retinotopic specificity is high, and for more global relationships to be established at higher hierarchical levels that could in turn reinforce the relations established at earlier levels. In this scheme, each set of grouped features in a scene would be defined by the formation of a cell assembly based on the synchronization of neuronal firing among populations of cells distributed within and across different levels of the cortical hierarchy. Different assemblies would be distinguished from one another by the independence of their firing patterns. Thus, multiple, highly salient distributed signals could coexist in the same network of cortical areas at the same time, each providing an independent representation of grouped features that could serve as the basis for attentional selection.

Why Has the Correlation Hypothesis Been So Attractive?

Although the temporal correlation model provides a number of fundamental advantages over competing theories, it remained an unpublicized theoretical curiosity until experimental evidence began to support some of its basic tenets. Numerous electrophysiological experiments had demonstrated that cortical neurons engage in synchronous firing on a millisecond time scale (Toyama et al., 1981a, 1981b; Michalski et al., 1983; Ts'o et al., 1986; Aiple and Krüger, 1988; Krüger and Aiple, 1988; Ts'o and Gilbert, 1988; Hata et al., 1991; Schwarz and Bolz, 1991). The majority of these studies, however, were conducted as a means for determining the functional connectivity among neurons. This approach was based on studies by Perkel and coworkers, who laid the foundations for interpreting cross-correlation measurements as physiological indications of anatomical connectivity (Perkel et al., 1967; Moore et al., 1970). Many cross-correlation studies were interpreted within this framework, as indicated by the lack of emphasis on the stimulus dependence of correlated activity and the use of pharmacological techniques to enhance neuronal firing rates in the absence of sensory stimulation.

A surge of interest in the correlation hypothesis took place when a series of experiments demonstrated that both short- and long-range synchronous firing in visual cortex were stimulus dependent (Gray and Singer, 1987, Soc. Neurosci., abstract, Gary and Singer, 1989; Eckhorn et al., 1988; Gray et al., 1989). These studies revealed that synchronous firing within and between columns of striate cortex was largely absent during spontaneous activity and could be evoked by appropriate visual stimulation. The correlated firing was accompanied by a rhythmic modulation of firing probability in the range of 30–60 Hz, suggesting that synchronous activity was an emergent phenomenon resulting from cooperative interactions in the cortical network. These findings were reminiscent of the theoretical predictions of Milner (1974), von der Malsburg (1981, 1985), and others (e.g., Braitenberg, 1978), because they demonstrated that synchronous firing was a transient, stimulus-dependent process. This interpretation stood in marked contrast to the more static notion that correlated firing was primarily the result of direct synaptic connectivity among cells. A further boost to the correlation theory came from the observation that synchronous firing depended not only on the presence of a stimulus but also on its global properties, such as common motion (Gray et al., 1989). Together, these findings led to a wave of new experimental studies and theoretical models (reviewed by Singer, 1993; Gray, 1994; Singer and Gray, 1995), although the latter far outnumbered the former.

What prompted this exuberance (some might say "irrational exuberance") of experimental and theoretical activity? I believe it can be traced back to widespread dissatisfaction with the conventional paradigm in sensory physiology. For roughly 40 years, much of sensory physiology was based on a common experimental paradigm and the results of experiments were often interpreted within a somewhat narrow theoretical framework. With few exceptions, measurements of neuronal activity were made one cell at time, and the variability of sensory responses, thought to reflect noise in the system, was removed by averaging, resulting in the conventional peri-stimulus time histogram or PSTH. This provided an accurate measure of a cell's mean firing rate as a function of time but also had the effect of removing much of the evidence of temporal structure in the neuronal activity. As a result, time, and hence neuronal dynamics, was relegated to a back seat in the focus of many researchers. The other major factor was the search for the ever-elusive representation of perceptual features and objects. This search was driven by the implicit assumption that the neural basis of sensory representation could be elucidated by mapping the receptive field properties of cortical neurons at different levels of the cortical hierarchy. If one knew the relationship between mean firing rate and stimulus properties, this would be sufficient to account for sensory representation. The rationale of this approach was clearly articulated by Barlow (1972) in his influential article entitled "Single units and cognition: a neurone doctrine for perceptual psychology." This approach led to extensive advances in our understanding of the cortical hierarchy but failed to provide an adequate theoretical framework for evaluating neuronal population dynamics. As a result, the dynamic behavior of neuronal populations had to be inferred from the mean firing rates and receptive field properties of single neurons measured one at a time.

The fundamental attraction of the temporal correlation model was that it provided a potential solution to the binding and superposition problems, while retaining all of the advantages inherent in population coding models.

It provided a physiologically plausible theoretical framework to account for the integration of distributed features into an emergent representation, and for distinguishing one representation from another under conditions when they would be present simultaneously. Moreover, it was envisaged as a dynamic process that depends on the properties of the visual features in an image. In this way, assemblies, defined by patterns of synchronous firing, could integrate features flexibly and with large combinatorial capacity. The initial evidence supporting this model unleashed a flurry of experimental and theoretical studies that were not limited to the visual system or notions of perceptual grouping (see Singer, 1999b [this issue of *Neuron*]). These were driven by the concepts of dynamic neuronal assemblies in which sensory, cognitive, and motor events are represented as dynamic patterns of activity in distributed populations of neurons (Vaadia and Aertsen, 1992; Bressler et al., 1993; Abeles et al., 1994; Varela, 1995; Fuji et al., 1996).

Evidence for and against the Correlation Hypothesis

An obvious question has arisen in the years since the initial findings supporting the correlation model: does the current body of evidence favor support for or rejection of the hypothesis? Alternatively, perhaps the key experimental tests of the hypothesis have yet to be done? In order to fully address these questions, it is necessary to briefly review the predictions of the theory. Because I have narrowed the scope of the original theories put forth by Milner (1974) and von der Malsburg (1981, 1985), I will limit my review to those predictions relating to preattentive perceptual grouping, first discussing the evidence that supports the predictions, and following with a discussion of evidence that would lead one to reject the hypothesis.

Supporting Evidence

Occurrence. The first and most obvious requirement of the correlation model is that synchronous neuronal firing should occur in the nervous system and in the visual cortex in particular. The plethora of studies over the past 20 years should leave no doubt that this is the case for the visual cortex, and for the nervous system in general (reviewed by Gray, 1994; Singer and Gray, 1995; Usrey and Reid, 1999; Singer, 1999b). Synchronous neuronal activity occurs over multiple temporal and spatial scales in the visual system. It can occur with high temporal precision where coincident spikes occur within 1 ms of each other. This type of synchrony, thought to reflect direct synaptic interactions, is most commonly observed between cells that are in close spatial proximity or that receive direct afferent connections (Toyama et al., 1981a; Usrey et al., 1999). Synchronous firing is also commonly observed at an intermediate time scale in which coincident spikes occur within ±10 ms of one another. This type of correlated activity occurs most often in response to sensory stimulation, can be sustained over periods ranging from tens to thousands of milliseconds, and can occur between cells located in the same cortical column, in different columns within the same cortical area, and in different cortical areas, as well as between cells located in the two cerebral hemispheres (Engel et al., 1991b; Singer and Gray, 1995). Because of its properties, this intermediate time

scale of synchronous firing is thought to provide a suitable, but not necessarily the only, substrate for the formation of dynamic cell assemblies. A third form of synchronous firing is commonly observed in which loosely coincident spikes occur within time scales of tens to hundreds of milliseconds of one another (Nelson et al., 1992; Nowak et al., 1995b, 1999). This type of correlated activity, also referred to as rate coherence (Vaadia and Aertsen, 1992), can occur spontaneously or be driven by visual stimulation, and has been observed over spatial scales similar to the intermediate form of synchrony. Taken together, then, the accumulated evidence indicates an abundance of synchronous activity in the visual system.

Proximity. Within the context of perceptual grouping, the correlation hypothesis makes a number of specific predictions regarding the properties of synchronous activity. The first of these predicts that synchronous firing should occur when cells are activated by stimulus features that exhibit one or more of the Gestalt grouping properties. This is a strong set of predictions, and each must be considered in turn. Gestalt theory predicts a strong association between features that lie in close proximity. In the context of the correlation model, synchronous firing should therefore fall off with distance in retinotopically organized cortical areas. The accumulated evidence supports this prediction, but the issue has been studied primarily in striate cortex. Here, the results are clear. The incidence and magnitude of synchronous activity drops off with the distance between cells, the largest separation showing positive correlation being ~10 mm (Michalski et al., 1983; Ts'o et al., 1986; Ts'o and Gilbert, 1988; Gray et al., 1989; Engel et al., 1990; Schwarz and Bolz, 1991; Livingstone, 1996). These intracortical distances within striate cortex typically correspond to visual field separations of less than 5°. Therefore, it will be interesting to see what type of relationship is present in extrastriate areas of the visual cortex, where receptive fields are larger and the spatial resolution is coarser.

Similarity. Another strong prediction of the correlation model is that synchronous activity should occur preferentially when cells are activated by features having similar properties. This should hold for many different feature categories such as orientation, color, and common motion. Here, again, these properties have been investigated primarily in striate cortex, and the bulk of the data relate to the orientation dependence of the interactions. Less information is available on common motion, and even less with respect to color (Ts'o and Gilbert, 1988). With respect to orientation, a number of studies have demonstrated that synchronous firing among cells with nonoverlapping receptive fields tends to occur most often when those cells have similar orientation preferences (Ts'o et al., 1986; Ts'o and Gilbert, 1988; Eckhorn et al., 1988; Gray et al., 1989; Schwarz and Bolz, 1991; Livingstone, 1996), although exceptions to this trend have been observed (Gray et al., 1989; Engel et al., 1990). This relationship is generally thought to reflect the bias for similar orientations in the patterns of horizontal intracortical connectivity in striate cortex (Gilbert and Weisel, 1983, 1989; Blasdel et al., 1992, Soc. Neurosci., abstract; Malach et al., 1993; Levitt et al., 1996; Bosking et al., 1997; Kisvarday et al., 1997; Schmidt et al., 1997). When

the separation between cells is smaller, such that they have overlapping receptive fields (i.e., <2 mm), the relation between orientation preference and synchronous firing is less clear. Correlated firing occurs with roughly equal probability for all combinations of orientation preference (Gray et al., 1989; Engel et al., 1990; Das and Gilbert, 1999). In the context of the correlation model, this may provide a mechanism for linking contours of different orientation, when the cues for spatial proximity dominate.

Common Motion. There have been fewer studies investigating the influence of common motion on synchronous firing, but they tend to confirm the predictions of the correlation model. A prominent effect of common motion was demonstrated in striate cortex by Gray et al. (1989). In two cases, they showed that pairs of cells separated by 7 mm engaged in synchronous firing when activated by bars drifting in the same direction and with the same velocity. This effect was absent when the same two bars drifted in opposite directions, in spite of the fact that both sets of stimuli elicited responses of roughly equal magnitude. A similar effect was demonstrated by Engel et al. (1991a) for cells recorded simultaneously in striate cortex and a motion-sensitive area of extrastriate cortex in the cat. In both of these studies, however, the samples were very small and the effects were measured using recordings of multiunit activity. This raised two concerns: the effect might be rare, and the differences in correlation strength between the two stimulus conditions might stem from the recruitment of different groups of cells during the responses. These criticisms were addressed in a study by Freiwald et al. (1995), who recorded from pairs of single units and demonstrated a similar effect of common motion on synchronous firing. The general trend of these findings was further confirmed by two later studies conducted in the monkey. Livingstone (1996) showed that synchronous firing was greater in magnitude when the stimuli activating the cells moved in the same rather than in opposite directions. Kreiter and Singer (1996) demonstrated an effect of common motion in the middle temporal area (MT) in the alert monkey. They recorded from pairs of cells having partially overlapping or nonoverlapping receptive fields but different direction preferences. When the cells were activated by two bars presented together, each moving over the receptive fields in the preferred direction of the cells, they observed little or no response synchronization. When the same cells were activated by a single contour, moving in a direction intermediate to that preferred by the two cells, they observed strong synchronization. This study provided convincing evidence for the dependence of synchronous activity on coherent motion. Two of the methods employed in this study left room for interpretation, however. The authors recorded multiunit activity in the majority of their measurements, and they selectively sampled only those pairs of recordings that exhibited synchronous activity to at least one stimulus condition. Thus, in spite of a strong effect, it is uncertain what percentage of cells within MT participate in this phenomenon.

Continuity and Segmentation. Aside from the issues of proximity and similarity, the correlation model also makes strong predictions regarding figural continuity, closure, and segmentation. Each of these properties provides strong cues for perceptual grouping. For example, when features, such as oriented contours, belong to a contiguous figure or one that is bounded on all sides, the model predicts that activity evoked by those features should be synchronous, even if they differ in orientation. Similarly, if more than one contiguous or closed figure is present in an image, the patterns of synchronous firing evoked by each figure should be independent of one another. These predictions have not been extensively tested, but a few experiments deserve mention. In the studies of Gray et al. (1989) and Freiwald et al. (1995) mentioned above, the strongest synchronization was observed when the cells were activated by continuous collinear contours. In a separate study of segmentation, Engel et al. (1991c) demonstrated that cells in striate cortex form separate synchronous assemblies when two drifting bars are presented together but move in orthogonal directions over the same region of the visual field. This study provided support for a key prediction of the correlation model. However, all of the measurements were made using multiunit activity, and the different stimuli used in the study led to marked differences in the firing rates of the cells. This raised the concern that the observed changes in synchronization might be due to the recruitment of different populations of cells as the stimulus was changed. Taken together, the results of these three studies support a role for synchronization in contour integration and feature segmentation, but further work is necessary to rigorously test the predictions.

Correlates with Perception. The other major prediction of the correlation model is that the patterns of synchronous firing evoked by perceptually grouped stimuli should correlate with an animal's performance on a perceptual discrimination task. Unfortunately, direct tests of this fundamental prediction have been few and limited in scope (de Oliveira et al., 1997; Fries et al., 1997; Lamme and Spekreijse, 1999). Thus, at present, the most important predictions of the theory are the ones for which the least experimental evidence is available. However, convincing evidence has been obtained for an indirect correlation between response synchronization and behavioral responses during interocular rivalry (Fries et al., 1997). In this study, recordings of multiunit activity were made from areas 17 and 18 in alert, strabismic cats while the animals viewed drifting gratings presented to either one or both eyes. The strabismus, induced at 3 weeks of age, shifted the distribution of ocular dominance in the adult animals so that the cortical neurons were only responsive to either the dominant or the nondominant eye. During binocular stimulation, each eye viewed a grating of the same orientation but drifting in opposite directions. This led to a rivalrous condition in which the eye mediating detection of the stimulus, usually the dominant eye, controlled the direction of optokinetic nystagmus (OKN). Making use of this effect, Fries et al. (1997) showed that synchronous activity, during rivalrous stimulation, was enhanced among the neurons supporting OKN and reduced among the neurons that were not supporting OKN. This effect occurred with no detectable differences in firing rates across the stimulus conditions they employed. This result provides the interesting, if somewhat complex, demonstration that the magnitude of synchronous firing, and not mean neuronal
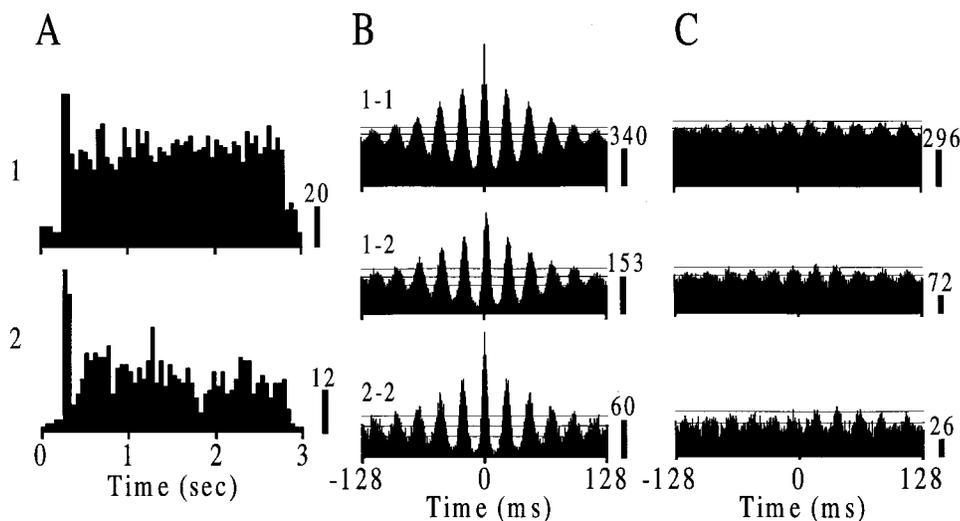
Figure 5. Stimulus-Evoked, Synchronized Oscillatory Activity in Striate Cortex of an Alert Macaque Monkey

(A) PSTH of multiunit activity recorded on two adjacent electrodes (1 and 2) separated by 300 μm. The cells had overlapping receptive fields and preferred the same orientation. They were coactivated by a single drifting grating of optimal orientation and direction.
(B) Auto- (1–1 and 2–2) and cross-correlation (1–2) histograms computed from the two spike trains during the response to the stimulus.
(C) Auto- and cross-correlation histograms computed from the same data after shuffling the spike trains by one stimulus period. The diminished amplitudes of the peaks in these histograms demonstrates that the synchronous rhythmic firing is not time-locked to the stimulus.

firing rate, correlates with the animal's perception of the stimuli.

While the results of this study provide support for the correlation hypothesis, several points deserve further discussion. First, the measured relations between perception and synchronization were indirect. OKN was not assessed during the periods of data collection, and thus the animal's perception of the stimuli could only be inferred from the previous behavioral tests. Similarly, while the eye that supports perception during rivalry determines the direction of OKN, this assay of behavior is limited in that it does not allow for a direct measure of perceptual discrimination. Finally, the experiments provide support for the hypothesis under conditions in which the normal physiology and anatomy of the cortex are perturbed (Löwel and Singer, 1992). It would be useful to determine if similar effects occur in animals with normal binocular vision.

Contradictory Evidence
Experimental evidence favoring a rejection of the correlation hypothesis has come from a variety of sources, but it has been more limited in scope than the supporting evidence. Critics of the hypothesis have largely based their arguments not on the validity of the contradictory evidence, but rather on criticisms of the supporting evidence and on arguments of the implausibility of the model. Here, I will focus on three sources of experimental evidence that contradict the claims of the correlation model. I will discuss the criticisms of the supporting evidence and issues of implausibility later.

Incidence and Properties of Neuronal Oscillations. When the experimental evidence supporting the correlation hypothesis was first reported, considerable emphasis was placed on the finding that synchronous firing in striate cortex often occurred during prominent oscillations of firing probability in the γ (30–60 Hz) frequency band (Gray and Singer, 1987, Soc. Neurosci., abstract;

Gray and Singer, 1989; Eckhorn et al., 1988; Gray et al., 1989, 1990; Engel et al., 1990). This led to a number of studies on the properties of stimulus-evoked γ band activity. Implicit in these studies was the notion that much could be learned about synchronization by elucidating the properties of oscillatory activity and its stimulus dependence. Several published studies favored rejection of the hypothesis on the basis that oscillatory activity was either not evident or did not exhibit the requisite stimulus dependence. Tovee and Rolls (1992a) reported that neuronal activity in the inferotemporal (IT) cortex of alert monkeys showed no signs of γ band oscillations when the cells were activated by stationary stimuli. Because IT cortex is considered to play a prominent role in object recognition, and thus feature integration, they argued that the absence of γ band activity favored the rejection of a synchronization-based mechanism for feature integration (Tovee and Rolls, 1992b; see also Engel et al., 1992a). A similar argument was put forth by Young et al. (1992) (see also discussion in Shastri and Ajjanagadde, 1993) on the basis of measurements in striate cortex and area MT of anesthetized monkeys and area IT of alert monkeys. They found little or no evidence for γ band activity in any of the three areas, even though they made explicit efforts to replicate the methods of previously published observations (Engel et al., 1990). They argued that γ band activity and, by extension, synchronous activity were phenomena limited to the cat and should therefore not be candidates for a general mechanism contributing to mammalian visual processing. This argument may have held weight, given the demonstration that long-range, stimulus-dependent synchronization in striate cortex of the cat is closely linked to rhythmic firing in the γ frequency band (König et al., 1995). But it was demonstrated in several later studies that synchronous γ band activity is a robust property of neuronal responses in areas V1

A

B

25

77

0   1   2   3
Time (sec)

0   20   40   60   80   100
Time (msec)

C

D

.2 nA

1

2

.8 nA

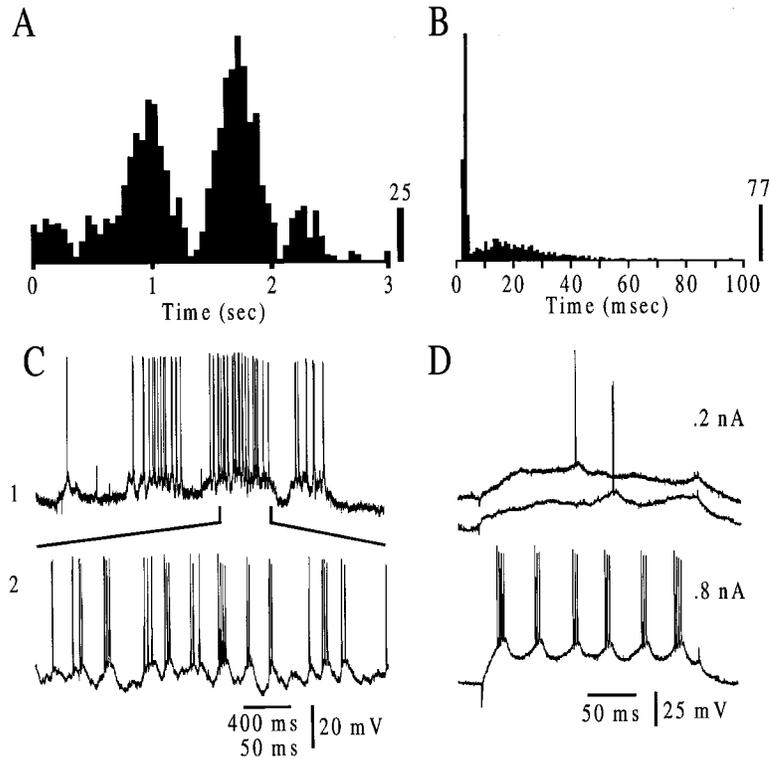400 ms | 20 mV
50 ms

50 ms | 25 mV

Figure 6. Activity Recorded from an Intrinsically Oscillatory (Chattering) Cell in Cat Striate Cortex in Response to Visual Stimulation and Intracellular Depolarizing Current Pulses

(A) PSTH recorded in response to 20 presentations of an optimally oriented, drifting light bar presented to the contralateral eye.
(B) The interspike-interval histogram (ISIH) computed from the spike trains sampled during the visual response. Note the characteristic bimodal shape of the distribution reflecting the intervals occurring within and between bursts.
(C) Membrane potential of the cell recorded during a single presentation of the stimulus. Trace 1 shows the complete 3 s trial. Trace 2 shows an expanded view of the activity during the peak of the visual response.
(D) Responses of the cell to two different levels of depolarizing current pulses. The cell shows no evidence of membrane potential oscillations when depolarized near the firing threshold (upper two traces). When the current intensity is increased to depolarize the cell well beyond firing threshold, a pronounced membrane potential oscillation is evoked, having a frequency near 30 Hz. Adapted from Gray and McCormick (1996).

and V2 of the alert (Figure 5), as well as anesthetized, monkey (Eckhorn et al., 1993; Frien et al., 1994; Friedman-Hill et al., 1995, Soc. Neurosci., abstract; Gray et al., 1995, Soc. Neurosci., abstract; Livingstone, 1996).

In spite of these later observations, there continues to be substantial debate about the prevalence and magnitude of γ band oscillatory activity and neuronal synchronization in the primate visual cortex (for examples, see Kreiter and Singer, 1992, 1996; Bair et al., 1994). Much of this debate is based on anecdotal reports at conferences or on unpublished observations alluded to in the literature. It seems reasonable to suppose that these divergent viewpoints may be compatible, if one considers the possibility that sampling biases and methodological differences contribute to the variations in results. Effects of this type may be particularly acute in the striate cortex, which contains many different cell types and a complex laminar and tangential organization. Similarly, differences in visual stimulation or analysis methods may lead to selective enhancement or suppression of rhythmic and/or synchronous activity (Gray and Viana Di Prisco, 1997). Each of these factors must be taken into account before firm conclusions can be reached.

In another set of experiments, Ghose and Freeman (1992) argued that synchronization-based grouping in cat striate cortex is not plausible, because γ band oscillatory activity is not stimulus dependent. Two lines of evidence figured prominently in their conclusion. They showed that oscillatory firing in striate cortex was strongest at threshold levels of stimulus contrast, indicating that it was occurring spontaneously and did not depend on visual stimulation. This line of reasoning was supported by separate measurements made in the lateral geniculate nucleus (LGN) of the thalamus. They

found pronounced γ band oscillations of activity in the LGN that occurred spontaneously and were in some instances suppressed by visual stimulation. From this, they concluded that cortical oscillations were driven by spurious correlations of spontaneous oscillatory activity in the LGN (Ghose and Freeman, 1992, 1997). Thus, not only was cortical oscillatory activity thought to be independent of visual stimulation, it might also be generated primarily by inputs from the LGN.

Several lines of evidence, however, argue for an alternative interpretation of these findings. Gray and Viana Di Prisco (1997) recorded neuronal activity in striate cortex of alert cats and found that oscillatory firing was stimulus dependent and rarely present in the absence of visual stimulation. They applied the analysis methods used by Ghose and Freeman (1992) and found that these techniques led to an artifactually high incidence of oscillatory activity. In another set of experiments, Gray and McCormick (1996) identified a population of neurons in cat striate cortex that exhibit oscillatory firing patterns generated by intrinsic cellular mechanisms (Figure 6). This finding argues for an intracortical contribution to the generation of γ band activity. Finally, several studies have now demonstrated that the robust oscillatory activity observed in the LGN by Ghose and Freeman is driven by retinal input (Laufer and Verzeano, 1967; Neuenschwander and Singer, 1995; Neuenschwander et al., 1999), occurs in a frequency range that is significantly higher than that commonly observed in striate cortex (Ito et al., 1994, Soc. Neurosci., abstract; Neuenschwander and Singer, 1995), and is only transiently synchronized to cortical activity in area 18 (Castelo-Branco et al., 1998; Herculano et al., 1999). Thus, contrary to the arguments put forth by Ghose and Freeman (1992, 1997), these findings indicate that cortical γ band activity arises

predominantly through intracortical interactions and is not driven by the higher-frequency oscillations generated in the retina.

A general point should also be made regarding the studies considered in this section: although studies of this type are useful in elucidating the properties and mechanisms of oscillatory activity, they are not sufficient to address the fundamental predictions of the correlation hypothesis. For this, experiments must be designed that rely on the measurement of neuronal activity at more than one site simultaneously. In this way, direct measurements of synchronization can be made without having to speculate about the properties of synchrony from single-unit recordings (Engel et al., 1992a).

Synchronization in a Motion Discrimination Task. Another source of evidence arguing against the correlation hypothesis has come from a recent study of correlated activity in extrastriate cortex (de Oliveira et al., 1997). These authors measured the activity of small groups of neurons, located within 900 μm of one another, in MT and the medial superior temporal area (MST) of monkeys, while the animals performed a simple direction discrimination task. The animals had to identify the direction of motion of a grating stimulus presented at various levels of contrast. Neuronal activity was recorded before and during the presentation of the stimuli, and cross-correlograms were calculated between all combinations of recorded cells. Interestingly, a high incidence of synchronous firing was found prior to the presentation of the stimulus. When the stimulus appeared, the firing rates of the cells increased, but the synchronization of the activity was reduced in a contrast-dependent manner. High-contrast stimuli that were easily discriminated resulted in a complete suppression of correlated firing. It was concluded that synchronous activity does not convey information about stimulus direction and therefore does not contribute to the direction discrimination task.

These conclusions appear sound and the evidence convincing, but there are also striking differences with earlier studies that need to be resolved. In contrast to the results of de Oliveira et al. (1997), Kreiter and Singer (1992, 1996) demonstrated that synchronous activity in area MT is visually driven and depends on the properties of the stimuli. They used drifting bars rather than gratings, but their animals were not engaged in a discrimination task and were only required to maintain visual fixation. This suggests that the attentive state of the animals in the study of de Oliveira et al. (1997) provides the most likely explanation for the different results. It is striking that attentive expectation might produce such a clear difference, particularly given the large body of evidence demonstrating the stimulus dependence of synchronous activity in other visual cortical areas (Singer and Gray, 1995; Usrey and Reid, 1999). However, neuronal synchronization has been shown to correlate with attentive behavior in other sensory and motor systems (Rougeul et al., 1979; Bouyer et al., 1981; Sanes and Donoghue, 1993; Murthy and Fetz, 1996a, 1996b). Therefore, the results of de Oliveira et al. (1997) must be taken seriously as evidence contrary to the correlation hypothesis. Further studies will be required to resolve the differences.

A Direct Test of the Correlation Hypothesis. Although many arguments and claims have been put forth to reject or support the correlation hypothesis, only one published study has carried out a direct test of several of the key predictions of the model (Lamme and Spekreijse, 1999). In order to place the results of this experiment in context, it is worthwhile to briefly review these predictions. According to the arguments put forth here and elsewhere (Milner, 1974; von der Malsburg, 1981, 1985; Singer, 1993; Singer and Gray, 1995; Roelfsema et al., 1996), the perceptual grouping of features in an image should be directly correlated with the probability and strength of synchronous firing of the neurons responding to those features. The activity of cells responding to other features should be temporally independent of the synchronously firing ensemble. For these predictions to be tested rigorously, it is necessary that the correlation measurements be made in an alert animal while it is performing a perceptual grouping task.

Lamme and Spekreijse (1999) recorded multiunit activity from chronically implanted electrodes while the animals performed a simple texture segregation task. The stimuli consisted of texture patterns composed of dense, uniformly oriented line elements or moving random dot patterns. The figures, defined by a difference in orientation or motion direction with respect to the background, were highly salient. Once the receptive field positions were mapped for a given pair of recordings, the stimuli were constructed to enable the testing of three predictions. On one set of trials, the figure was positioned over both receptive fields of a pair. In this condition, the cells should synchronize their activity. On a second set of trials, the figure was positioned to cover one receptive field, while the other receptive field was covered by the background. In this condition, no synchronization should be observed. And on a third set of trials, the figure was positioned so that both fields were covered by parts of the background. The prediction of the correlation hypothesis is less clear in this case, but given that the background pattern was highly coherent, one might expect to see correlated activity.

The results of these experiments were fairly clear. There was no significant difference in the magnitude of synchronization across the three stimulus conditions when the population of measurements was compared, leading the authors to conclude that response synchronization does not correlate with texture segregation. This seems like a reasonable interpretation of the data and may be correct. However, a closer look at the details of the experimental design and the data analysis reveals another story. There are three key points that need to be made. First, the local properties of the stimuli activating the cells were changed substantially across stimulus conditions. When the figure was presented over both receptive fields, each cell was stimulated by the same orientation or motion direction. But when the figure was shifted in position, so that it covered one field and the background covered the other field, the local orientation or direction of motion stimulating one of the cells was now different by either 20° or 90° for the oriented texture, or 180° for the drifting dots. Although an attempt was made to counterbalance this effect, this is clearly a major change for neurons in V1 that are known to be selective for stimulus orientation and direction of motion. Second,

the neuronal recordings were composed of broadly tuned multiunit signals obtained from chronically implanted electrodes. The neuronal responses displayed very weak orientation selectivity, suggesting that signals were being sampled from a comparatively large volume of cortical tissue. This means that when the local stimulus orientation or motion direction was changed, the signals at each site continued to show vigorous responses because different sets of neurons were being recruited. A condition of this sort makes interpretation of the cellular interactions difficult at best, because the population of active cells is changing as the stimulus changes. Finally, it is doubtful whether the neuronal activity displayed synchronous firing under any of the stimulus conditions. The published cross-correlograms exhibit relatively broad peaks (20–50 ms) centered at zero time lag, an effect that is consistent with synchronized activity at a time scale in the tens of milliseconds. But if one inspects the PSTHs, it is clear that the cells respond to the stimuli with an initial, transient burst of activity followed by a much lower and sustained level of firing throughout the trials. This type of covariation in the rate and latency of neuronal responses has recently been demonstrated to lead to correlation peaks of the type just described (Brody, 1999a, 1999b). Thus, it is possible that the published cross-correlations in firing do not reflect true synchronization, but a common temporal envelope in the cellular responses. Needless to say, a result of this type would also provide evidence contrary to the theory but of a less convincing nature than that claimed by the authors.

Given these considerations, what conclusions can be reached from this experiment? First, the results do not provide a basis for rejecting the correlation hypothesis. They provide limited negative evidence for a phenomenon that is intrinsically difficult to observe. Second, the study reveals certain aspects of an experimental design that are necessary for a rigorous test of the hypothesis. The local properties of the stimulus should remain relatively constant, in order to maintain comparable firing rates, while changes in the relations among features should yield distinct perceptual differences. When possible, recordings should be made of single-unit activity to facilitate the interpretation of the data. And appropriate analytical tools should be applied to rule out spurious sources of correlation when the neuronal responses themselves are nonstationary.

Why Has the Correlation Hypothesis Been So Controversial?
At the time of its inception, the correlation hypothesis put forth by Milner (1974) and von der Malsburg (1981, 1985) was based largely on theoretical considerations, with very little reliance on experimental evidence to support its claims. Drawing on the concepts of Hebbian cell assemblies (Hebb, 1949; Braitenberg, 1978), the model postulated a mechanism to solve several inherently difficult problems originating from perceptual psychology (Neisser, 1967) and theories of representation in artificial networks (Minsky, 1961; Arbib, 1964). In this context, the theory remained largely uncontroversial, but didn't gather much attention either. The status of the theory changed dramatically when experimental evidence provided support for some of its basic tenets (Gray and

Singer, 1987, Soc. Neurosci., abstract; Gray and Singer, 1989; Eckhorn et al., 1988; Gray et al., 1989, 1990; Engel et al., 1990), whereupon it was rapidly embraced by some and hotly disputed by others. What factors spurred the resulting debate, which has continued unabated into the present? I believe that the origins and current basis of the controversy can be traced to four sources.

Interpretation, Emphasis, and Speculation
Following the early experimental studies claiming support for the correlation hypothesis, several things happened which shaped the debate of the issues. First, the robust, stimulus-dependent synchronization of activity within (Gray and Singer, 1987, Soc. Neurosci., abstract; Gray and Singer, 1989; Gray et al., 1989; Engel et al., 1990) and between (Eckhorn et al., 1988) cortical areas was interpreted as evidence for a mechanism underlying perceptual grouping in the visual system. This bold claim received widespread attention from experimentalists and theoreticians alike (Kammen et al., 1989; Sporns et al., 1989; Stryker, 1989; Baldi and Meir, 1990; Schuster and Wagner, 1990; Sompolinsky et al., 1990). Others argued that the results were preliminary—perhaps even interesting and deserving of further study, but not yet warranting such claims. Second, the early studies revealed a phenomenon new to the visual cortex, stimulus-evoked high-frequency oscillations of firing. This phenomenon was attributed with particular significance as a mechanism to facilitate synchronization (Eckhorn et al., 1988; Gray et al., 1989), which led to a large number of theoretical investigations involving the analysis of networks of coupled oscillators (Baldi and Meir, 1990; Niebur et al., 1990; Schuster and Wagner, 1990; Sompolinsky et al., 1990). The focus on oscillations naturally led to an emphasis on phase and frequency as coding parameters, which had its own set of implications. Now the interactions between neuronal populations could potentially be reduced to a couple of parameters, making it mathematically attractive. This in turn raised questions as to how many assemblies could coexist in the same network given the limited number of frequencies and phases (Lisman and Idiart, 1995); it even began to appear as though some might take the analogy to the FM radio seriously (Hoppensteadt and Izhikevich, 1998). Third, two influential studies were published that claimed a role for synchronous oscillations in visual attention and conscious experience (Crick and Koch, 1990a, 1990b). In spite of their clearly speculative nature, these manuscripts attracted further widespread interest. Thus, from a limited number of observations, the correlation hypothesis was expanded far beyond its original scope. The speculative nature of the dialog no doubt raised the blood pressure of some.

Difficulties with Replication
Another source of controversy surrounding the correlation hypothesis stemmed from the reported difficulties in replicating some of the original experimental findings. As mentioned earlier, these reports either failed to find evidence of oscillatory activity in the primate visual system (Tovee and Rolls, 1992; Young et al., 1992; Bair et al., 1994), or reported oscillatory activity that was not stimulus dependent (Ghose and Freeman, 1992). In the years since, much of this debate has been cleared up

by the demonstration that synchronous oscillatory activity is stimulus dependent and readily apparent in the striate cortex of monkeys as well as cats (Eckhorn et al., 1993; Frien et al., 1994; Friedman-Hill et al., 1995, Soc. Neurosci., abstract; Gray et al., 1995, Soc. Neurosci., abstract; Livingstone, 1996; Fries et al., 1997; Gray and Viana Di Prisco, 1997). This has shifted the debate from whether these phenomena occur to whether they are of functional importance. However, anecdotal and unpublished reports continue to persist that oscillatory activity is absent or hard to find. Unless findings of this sort are published, these accounts are likely to have little impact other than to fuel the debate.

It is interesting to note that much of the debate surrounding the prevalence and properties of oscillatory activity had little to do with the original formulation of the correlation hypothesis. This debate owed its existence to the emphasis placed on oscillatory firing by the early experimental and theoretical studies. Although current formulations of the correlation hypothesis continue to favor a mechanistic role for oscillatory activity, measuring the properties of oscillatory firing one cell at a time is insufficient for testing the key predictions of the hypothesis. Synchronous activity exists in many forms and can only be assessed by recording from at least two cells simultaneously.

Physiological Plausibility

In recent years, a new challenge to the correlation hypothesis has been put forth. Powerful arguments have been raised on cellular and biophysical grounds to claim that a synchronization-based mechanism for solving the binding problem is physiologically implausible (Shadlen and Newsome, 1994, 1998; Shadlen and Movshon, 1999 [this issue of Neuron]). The essence of this argument is that cortical neurons themselves suffer from a binding problem of sorts (Shadlen and Newsome, 1998). A typical cortical neuron receives ~5000 synaptic inputs (Braitenberg and Schuz, 1991), ~80% of which are excitatory, and at least half, if not more, of these inputs come from other cortical neurons within a radius of ~200 μm. Because of columnar organization, these nearby neurons generally have similar functional properties. Thus, any stimulus that excites one neuron is likely to excite a large number of adjacent neurons that provide input to it. Visual cortical neurons excited by a sensory stimulus usually exhibit firing rates in the range of 10–100 spikes per second, and the efficacy of excitatory synaptic inputs onto cortical neurons ranges from 0.1 mV to 1.0 mV (Mason et al., 1991; Nicoll and Blakemore, 1993; Thomson and West, 1993; Matsumura et al., 1996; Thomson and Deuchars, 1997). These considerations raise a substantial problem when neurons operate within what Shadlen and Newsome (1998) call a "high input regime": how can a neuron engage in selective synchronous interactions with a subset of its inputs when a large percentage of all the cell's inputs are active and synchronous? According to Shadlen and Newsome (1994, 1998) (see also Shadlen and Movshon, 1999), this scenario presents an insurmountable problem for any mechanism that relies on a temporal patterning of spike activity, synchronous assemblies included.

At first glance this argument appears articulate, solid, and well grounded. In fact, it may even leave those with an interest in temporal correlation feeling as though they are pursuing the wrong questions. But closer scrutiny reveals a few weaknesses in the edifice. First of all, the widespread existence of many different forms of temporal patterning and synchronization in the nervous system is empirical proof that the obstacles posed by the "high input regime" are overcome, sometimes with surprising effectiveness (for examples, see Gray, 1994; Singer and Gray, 1995; Usrey and Reid, 1999; see also Singer, 1999b). But how could this be the case, if neurons don't possess the mechanisms to select synaptic inputs on the basis of their timing? Another large body of evidence appears to suggest otherwise. There are an abundance of cellular mechanisms that can contribute to the temporal structure of neuronal activity. Although the examples are numerous, it is worth considering a few of them in detail to make the point.

The first of these to consider is inhibition. Although Shadlen and Newsome (1994, 1998) treat inhibition as providing a balancing function to stabilize network activity, there is extensive evidence demonstrating that inhibition can also shape the temporal structure of neuronal activity on a millisecond time scale (Freeman, 1975; Buzsaki et al., 1983; Llinas et al., 1991; Lytton and Sejnowski, 1991; Wilson and Bower, 1992; van Vreeswijk et al., 1994; Bragin et al., 1995; Cobb et al., 1995; Whittington et al., 1995; Ylinen et al., 1995; Bush and Sejnowski, 1996; Plenz and Kitai, 1996; Traub et al., 1996; Wang and Buzsaki, 1996; Chrobak and Buzsaki, 1998; Pauluis et al., 1999). One consequence of inhibition is that excitatory synaptic inputs arriving during its occurrence are less efficacious in activating the postsynaptic cell because of the shunting and hyperpolarizing effects of the inhibitory potentials.

Intrinsic voltage-gated conductances provide another class of mechanism capable of shaping the temporal structure of neuronal activity (Llinas, 1988). It is now well established that cortical neurons exhibit a variety of distinct firing patterns in response to depolarizing inputs (Connors et al., 1982; McCormick et al., 1985; Connors and Gutnick, 1990; Gray and McCormick, 1996; Steriade et al., 1998). These patterns range from tonic firing with varying rates of accomodation, to the generation of bursts and intrinsic oscillations. A variety of voltage-gated, dendritic, and somatic currents underlie these phenomena (Johnston et al., 1996; Yuste and Tank, 1996), including outward K^+ currents that result in afterhyperpolarization or pacemaker activity (Storm, 1990; Hoffman et al., 1997; Luthi and McCormick, 1998) and inward Na^+ and Ca^{2+} currents that can amplify synaptic inputs (Schwindt and Crill, 1995; Stuart and Sakmann, 1995; Margulis and Tang, 1998), produce rebound excitation, or evoke Ca^{2+}-dependent action potentials (Schiller et al., 1997). Each of these currents exhibits distinct voltage- and time-dependent properties; thus, their contribution to the membrane potential of a cell will depend not only on the instantaneous voltage but also on the time course of the voltage change.

Some of these conductance mechanisms can have profound influences on the excitability of cortical neurons by enhancing synaptic inputs in a time-dependent manner. For example, Larkum et al. (1999) have recently shown that excitatory dendritic synaptic input, coincident within a few milliseconds of back-propagating Na^+ action potentials, facilitates the generation of dendritic

Ca^{2+} spikes. Margulis and Tang (1998) have shown that coincident excitatory dendritic input leads to supralinear summation that depends on voltage-gated Na$^+$ currents. Volgushev et al. (1998) demonstrated that the sensitivity of cortical neurons to weak excitatory synaptic input depends precisely on the phase of subthreshold oscillations of membrane potential. Azouz and Gray (1998, Soc. Neurosci., abstract) have found that the action potential threshold in visual cortical neurons recorded in vivo is inversely correlated with the rate of membrane depolarization preceding a spike. These findings demonstrate that cortical neurons possess a variety of voltage-gated inward currents that act postsynaptically to amplify the response of neurons to synchronous synaptic inputs.

Other conductance mechanisms, such as those responsible for generating rapid burst discharges (Traub et al., 1991; Rhodes and Gray, 1994; Gray and McCormick, 1996), may act indirectly to facilitate synaptic transmission in a time-dependent manner. The rapid firing within a burst can lead to temporal summation postsynaptically (Miles and Wong, 1986) and transiently increase the probability of neurotransmitter release (Thomson et al., 1993; Stevens and Wang, 1995). Such a mechanism ensures that bursts, rather than single spikes, will be more effective in mediating synaptic transmission and thus driving the postsynaptic neuron (Gray and McCormick, 1996; Lisman, 1997; Wang, 1999). Moreover, when burst firing occurs synchronously in a population of cells, the efficacy of the resulting synaptic input is likely to be much greater than that occurring among cells firing trains of uncorrelated spikes (Gray and Viana Di Prisco, 1997).

Together, these and other studies demonstrate that intrinsic voltage-gated conductances can regulate the time course and pattern of spike activity, contribute to the effectiveness of synaptic transmission, and endow neurons with an exquisite sensitivity to the timing of their synaptic inputs. In this context, it no longer seems reasonable to think of cortical neurons as integrate-and-fire devices that accumulate input over a time course that is determined by their passive membrane time constant. The time constant itself changes continuously with variations in membrane conductance resulting from synaptic input (Borg-Graham et al., 1996, 1998; Koch et al., 1996). Therefore, conclusions regarding the timing of neuronal activity that are based on such assumptions are likely to be wrong.

Paradigmatic Differences

On a final note, it should be pointed out that some hypotheses—and the correlation hypothesis is no exception—meet resistance because they represent a significant deviation from a widely accepted theoretical framework. As mentioned earlier, much of cortical physiology has been dominated by the view that cortical function can be elucidated by measuring the mean firing rates of neurons one cell at a time. Temporal fluctuations of activity are often considered to be noise obscuring a signal. This view provides a simple and tractable framework in which the problems of neural coding can be reduced to the counting of spikes. This is very attractive, given the difficulties associated with measuring the activity of even small groups of neurons. If, on the other hand, perceptual phenomena result from the dynamic and nonstationary interactions of large populations of

neurons, controlled by a myriad of nonlinear cellular mechanisms, many of the conventional ways of thinking about cortical organization and measuring its properties will have to change. This argument has no direct bearing on the validity of the correlation hypothesis; however, it may help to account for some of its controversial nature.

Future Experimental Tests of the Correlation Hypothesis

If the correlation hypothesis, as I have outlined it, is to survive as a viable model of a mechanism contributing to perceptual grouping and segmentation, it will have to withstand a rigorous test of its predictions. What sort of experiment could serve as an adequate test of the hypothesis? If the outcome of an appropriate experiment were positive, what form would the resulting evidence take? Put another way, what type of evidence would satisfy the critics of the theory? Perhaps equally important, what type of evidence would favor a definitive rejection of the hypothesis? This latter issue is difficult to address, because negative evidence is inherently problematic. One can easily resort to a variety of claims to discount it. Here, I will attempt to outline a specific set of criteria that an experimental paradigm must meet to provide a rigorous test of the hypothesis, as I have stated it. I will also discuss what form the evidence might take that would be consistent with a positive outcome for such an experiment.

One question to ask is where to look. There are many areas of visual cortex that no doubt play an important role in perceptual grouping. And good arguments can be made for a number of areas as the appropriate candidate. For several reasons, primary visual cortex is as good a place as any. First, it is well established that striate cortex exhibits robust, stimulus-dependent synchronous activity over a range of spatial and temporal scales. Second, anatomical studies have revealed a rich network of horizontal and feedback connections within striate cortex that could mediate grouping functions (Rockland and Lund, 1982; Gilbert and Weisel, 1983, 1989; Malach et al., 1993; Levitt et al., 1996; Bosking et al., 1997; Schmidt et al., 1997). And third, a variety of physiological studies in striate cortex have provided evidence for contextual effects related to perceptual grouping (Li and Li, 1994; Kapadia et al., 1995; Lamme, 1995; Polat et al., 1995; Sillito et al., 1995; Zipser et al., 1996; Levitt et al., 1997; Ito and Gilbert, 1999).

Another question to address is what type of stimulus and task should be employed. Here, the answer is fairly clear. The stimuli should be composed of a collection of features that pop out preattentively from a background on the basis of their Gestalt properties. Ideally, these features would contain multiple cues for grouping such as continuity, closure, and common motion. This would make them highly salient and easily detectable as a figure. The stimuli should be constructed so that the local features activating the cells can be held relatively constant, while the relations between the features can be modified to change the global properties of the image. This would be essential, so that the salience of a figure with respect to the background could be varied along a continuum, thereby varying the difficulty of the task. The task itself must require the active participation

of the animal, so that one could be relatively certain that the animal sees and responds to the figure. This could be attained by simply having the animal detect the presence of a figure by looking at it, or by performing a discrimination task. Having control over the salience of the figure would then enable the experimenter to evaluate the relation between neuronal activity (i.e., temporal correlations of activity) and behavioral performance. Finally, the measurements of neuronal activity must be made at the same time that the animal is performing the task. This is the only way to establish a convincing correlation between neuronal activity and behavior.

If such an experiment were conducted, what type of evidence would constitute convincing support for the correlation hypothesis? Here, the answer is straightforward in some sense, but it may be important to establish criteria that are not too rigid in order to accommodate unexpected observations. The simplest prediction would be as follows: two neurons should exhibit *synchronous firing* when they are activated by *different parts* of a *perceptually segmented* figure. When the position of the figure is changed, so that one of the neurons is activated by an element of the background and the other by an element of the figure, their activity should be temporally independent. In the former case, the strength and/or probability of synchronous firing should covary with the salience of the figure and with the behavioral performance of the animal.

While this prediction appears definitive, it is important to clarify a few of its key components. First, what is meant by synchronous firing? What are the appropriate time scales, latencies, and durations of synchronous activity that are consistent with the predictions? Much of the evidence from striate cortex suggests that synchronous activity should occur over a time scale of a few (e.g., ±5) milliseconds. It is not unreasonable to think, however, that slower time scales of correlated firing, on the order of 10–50 ms, may be the relevant parameter (Vaadia and Aertsen, 1992; Nowak et al., 1995b, 1999). Synchronous activity should develop rapidly after the onset of the stimulus, probably within 50–150 ms, and should certainly be apparent before the onset of a motor response. The duration of synchronous firing is more difficult to predict. It is conceivable that only one or a few synchronous spikes would be sufficient to signal the relations among features, or it may require a more sustained period of correlated activity lasting tens, or perhaps hundreds, of milliseconds. In any case, such brief periods of correlated firing are likely to be difficult to detect and will require analysis methods that are sensitive to such brief correlations (Aertsen et al., 1989; Gray et al., 1992; Vaadia et al., 1995; Riehle et al., 1997).

Second, what constitutes different parts of the same figure? Here, the prediction will vary with the cortical area being investigated. For striate cortex, particularly that of the primate, the spatial scale is likely to be critical. It does not seem reasonable to postulate correlated firing between cells that are separated by great distances and hence are responding to features that are widely separated in the visual field. The most appropriate scale to search for neuronal interactions in primate striate cortex is likely to fall within 5 mm of separation and thus a few degrees of visual angle. The figures will have to be constructed to be consistent with the predicted spatial scale of the interactions, and this will vary depending on which cortical area is being investigated.

Third, what properties of the stimulus define it as being perceptually segmented? This may be a particularly important point. The visual system is likely to signal figure–ground relationships by using the most efficient mechanism available to it. This means that firing rates across a population of cells may suffice in many instances (e.g., a simple high-contrast object against a uniform background). Given this consideration, it is possible that selective synchronization may come into play in those situations where ambiguities in the stimulus must be resolved (Leonards et al., 1996; Singer, 1999b). Thus, it will be important to use stimuli composed of similar feature elements in which the figure to be detected is defined by one or more of the Gestalt properties relating those elements to one another. Pushed to speculate on this point, I would argue that continuity, closure, and common motion are likely to be the most important.

Supposing that all of the above criteria were met and a result consistent with the predictions was obtained, what other criteria would have to be met for the evidence for the correlation hypothesis to be considered convincing? The answer here is that it should be a robust phenomenon that is present in a significant fraction of the cell population. The difficult question is how robust, and in what percentage? These are qualitative measures that depend on one's own definitions. But these are issues that will have to wait for the results of further experiments.

Finally, what type of negative evidence would favor definitive rejection of the correlation hypothesis as I have stated it? This is more difficult, because negative results from the experiments proposed above probably would not be sufficient to completely reject the hypothesis. One could always resort to the argument that the proposed mechanisms are implemented outside of striate cortex. However, it might be possible to rule out a given cortical area if the results met certain criteria. For example, if there were no detectable differences in the incidence or magnitude of correlated firing across stimulus conditions, this might provide a strong case. I have stated my reasons earlier for why I believe the experiment of Lamme and Spekreijse (1999) does not meet these criteria. Another scenario might be that correlated firing is conspicuously absent or reduced in strength for those stimuli fulfilling the Gestalt grouping criteria. This might indicate that the absence of correlated firing is the relevant signal (see de Oliveria et al., 1997, for a related result). Finally, it might be the case that correlated firing is absent under all of the stimulus conditions for a given experiment. This is likely to be less convincing, because it could be argued that the relevant cell populations were not sampled during the experiment. In the end, if the hypothesis is incorrect, it will require negative evidence from a variety of experiments to kill it. On the other hand, if it is correct, a few key positive results, replicated for rigor, might be sufficient for even the critics to accept it.

Conclusions

Having reviewed some of the available evidence for, as well as against, the temporal correlation hypothesis, it is appropriate to ask what conclusions can be drawn

with confidence and what questions remain open for future studies. As I see it, these questions fall into four general categories. The first concerns the psychophysical foundations of the theory. As I have outlined here, the correlation hypothesis relies on the notion that perceptual grouping is a rapid, preattentive process that acts in parallel over the visual field to "bind" features with common properties into larger units. Although there is much ongoing debate concerning the details of this process, nearly 80 years of psychophysical research indicate that this fundamental concept is sound.

Given this conclusion, the second category of questions concerns the neurophysiological basis of perceptual grouping. Because of the retinotopic organization of visual cortex and its parallel nature, grouping must involve the integration of information from distributed populations of cells. I have briefly reviewed the evidence for two candidate mechanisms, hierarchical convergence and population coding. The evidence indicates that both mechanisms operate together, but concerns raised by the correlation hypothesis suggest that additional mechanisms must be involved. In its simplest form, convergence coding cannot solve the problem, because the constraints on flexibility, anatomical wiring, and numbers of cells are too severe. Population coding appears to solve these problems, but may itself suffer from a superposition problem that is due to the parallel nature of the grouping process. Some mechanism should be available to distinguish one group of bound features, and hence one population of active neurons, from another. This argument appears valid, but it should be pointed out that it is primarily a theoretical construct. We don't know for certain that a "binding" problem exists that requires a neural solution. If we accept that superposition is a problem, then it appears that some form of assembly-based representation is required. The correlation hypothesis posits that cell assemblies are formed when a common pattern of synchronous firing is established among the participating cells.

This raises a third set of questions. Does existing evidence support this definition of cell assemblies, and if so, do they exhibit the properties required by the theory to account for perceptual grouping? The answer to the first question appears to be yes. There is an abundance of evidence, from visual cortex and many other neural structures, demonstrating the existence and properties of dynamic ensembles of synchronously firing cells. The answer to the second question is more equivocal. The evidence for synchronous activity clearly satisfies some of the requirements of the correlation hypothesis. But the strongest predictions of the theory are also the ones for which the least evidence is available. In spite of this, and in contrast to the viewpoints expressed by some critics of the theory, none of the existing evidence points to an outright rejection of the theory. The resolution of these issues will have to await the results of future experimental studies.

A fourth category of questions concerns the physiological plausibility of the theory. It has been argued that cortical neurons may not possess the requisite biophysical mechanisms to establish synchronous interactions with subsets of their active inputs. I have argued to the contrary that cortical neurons possess an impressive array of mechanisms to generate and respond to patterns of synchronous activity. It remains to be determined how these mechanisms operate together to select certain inputs for processing. In this context, the anatomical connectivity that provides the substrate for perceptual grouping is critical. The appropriate connections must be available to provide a wide repertoire of possible dynamic assemblies. In his early formulations of the theory, von der Malsburg (1981) proposed the existence of a network of horizontal connections in visual cortex that links populations of neurons with similar receptive field properties. In the years since, much evidence has accumulated to support this prediction (Gilbert and Weisel, 1983, 1989; Blasdel et al., 1992, Soc. Neurosci., abstract; Malach et al., 1993; Levitt et al., 1996; Yoshioka et al., 1996; Bosking et al., 1997; Kisvarday et al., 1997; Schmidt et al., 1997). When these data are combined with the finding that local (i.e., $<500 \mu m$ radius) intracortical connections are relatively homogeneous, the anatomical network appears to be well suited for linking activity across a wide array of features (Engel et al., 1990; Das and Gilbert, 1999). These findings suggest the interesting possibility that the Gestalt criteria for perceptual grouping reside in the network of intracortical connections (Roelfsema et al., 1996). If so, this would provide for a large number of possible interactions, with the stimuli and the underlying neuronal dynamics ultimately determining which interactions take place and which features become grouped (Sporns et al., 1991; Roelfsema et al., 1996; Yen and Finkel, 1998; Li, 1999a, 1999b; Yen et al., 1999).

In conclusion, the temporal correlation hypothesis remains an attractive theoretical framework for investigating the neural basis of perceptual grouping. The experiments I have proposed provide one way to test several of the key predictions of this theory, but these studies are still correlative in nature. A true test of the predictions would be one in which a causal link could be established between synchronous activity and perception. Recent experiments in the insect olfactory system demonstrate that tests of this sort can be accomplished in simpler nervous systems (Laurent, 1996; MacLeod and Laurent, 1996; Wehr and Laurent, 1996; Stopfer et al., 1997). A similar experiment would clearly be difficult to achieve in the primate visual system. Any manipulation that might perturb the patterns of synchronous firing would be likely to influence the overall distribution of activity as well. Consequently, a rigorous test of the correlation hypothesis may ultimately have to rely on a combination of theoretical and experimental evidence obtained from a variety of studies.

References

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