Synchrony Unbound: A Critical Evaluation of the Temporal Binding Hypothesis

Review

Michael N. Shadlen* and J. Anthony Movshon[†‡]
*Department of Physiology and Biophysics and
Regional Primate Research Center
University of Washington
Seattle, Washington 98195
[†]Howard Hughes Medical Institute and
Center for Neural Science
New York University
New York, New York 10003

Introduction

In the early stages of visual processing, objects and scenes are represented by neurons with small visual receptive fields. Each neuron provides information about local features of a scene, but to describe a scene in terms of objects requires that these features be combined. Objects can cover wide areas of visual space and be partially occluded by other objects, so the problem of binding the separate representations of parts into coherent wholes is not a simple one. This "binding problem" has received considerable attention. Gestalt psychologists articulated a number of principles for grouping and organizing scene elements (Köhler, 1930; Koffka, 1935; Kanisza, 1979), and more recently the related problem of image segmentation has received a good deal of attention in computer vision (see, for example, Pal and Pal, 1993, for a review). The binding problem is really best considered as a series of related problems, all of which require the combination of information from multiple sources. Information must be integrated across the visual field and combined according to specific attributes. Some of those attributes are pictorial features like line orientation, texture, color, simultaneity of appearance, and common motion, but others require more complex information about such things as 3D shape, lighting, and object surface properties (Gregory, 1970; Marr, 1982; Shimojo et al., 1989; Adelson, 1993; Ullman, 1996; Adelson, 1999; Kersten, 1999).

All of these levels of representation must be combined to solve the binding problem. A simple example is caricatured in Figure 1A. The four arrows are effortlessly perceived as separate overlaid objects, but the computations that generate this percept must be informed by notions of occlusion and object continuity and must remain unconfused by serendipitously shared features like orientation and surface color.

Higher-level vision poses many problems; some, like visual object recognition, seem at least as difficult to solve as the binding problem. But, in recent years, von der Malsburg and others have advanced the view that binding is a special problem and requires a special solution (von der Malsburg, 1981, 1985, 1995; von der Malsburg and Schneider, 1986; Reitboeck et al., 1987; Wang et al., 1990; Grossberg and Somers, 1991; Sporns et al., 1991; Neven and Aertsen, 1992; Tononi et al., 1992;

Singer, 1994). The basis for this argument is that it is necessary to "tag" each visual neuron to signify the object to which its activity relates. Each neuron therefore has to carry two distinct signals, one that indicates how effective a stimulus is falling on its receptive field, and a second that tags it as a member of a particular cell assembly. To make these signals distinct, von der Malsburg (1981) proposed that the "effectiveness" signal would be carried by a conventional rate code, while the "tag" signal would be created by synchronizing the spike activity of the neuron with spikes from other neurons in the same assembly. This novel idea has led to a great deal of experimental work and to several further elaborations of the original theory.

In this paper, we will articulate our doubts and concerns about this theory and its experimental support. We consider first whether the theory is an a priori reasonable approach to solving the binding problem, and conclude that it is at best incomplete. We then ask whether spike synchrony can plausibly be used as an informational code, and conclude that there are significant practical and theoretical obstacles both to encoding and to decoding information in this way. We then examine the experimental evidence usually adduced to support the synchrony hypothesis, and conclude that the evidence is largely indirect and has no proven relevance to the issue of binding per se. We will finish by asking whether the binding problem is truly of unique difficulty and requires a unique solution, and by considering some strategies for solving the binding problem that do not require the creation of a special neural code.

Addressing the Binding Problem with a Temporal Code
The Temporal Binding Hypothesis

Von der Malsburg (1981) proposed temporal correlation to escape a combinatorial problem in neural coding. Theories that propose the creation of "cardinal" cells to represent particular combinations of signals from lower-order neurons are implausible because the number of combinations to be coded exceeds the number of neurons available. In Von der Malsburg's theory, the activity of low-order neurons would be combined *only* when their spike activity was synchronized to within a few milliseconds to create a synchronously active cell assembly (Hebb, 1949; Braitenberg, 1978; Abeles, 1991). Synchronization would be dynamically modulated, so that a particular cell could belong to one cell assembly at one moment and to a second at another; in this way, the combinatorial bullet could be dodged and arbitrarily large numbers of states coded with a reasonable number of neurons.

Although he conceived the "temporal correlation theory" to have broad applications to neural computation, Von der Malsburg offered the specific problem of figure–ground discrimination as a sample case, suggesting that spike synchronization would group together the elements that make up figure and ground (see Lamme and Spekreijse, 1998, for a test of this particular idea). The most enduring application of this theory has been the

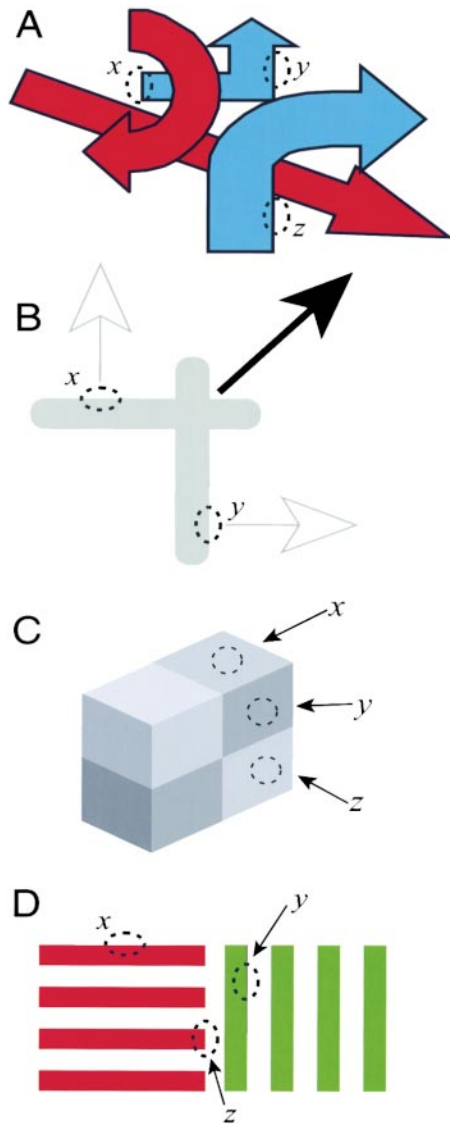[‡] To whom correspondence should be addressed (e-mail: movshon@nyu.edu).

Figure 1. Feature Binding and Related Problems in Visual Perception

(A) Contours falling in the receptive fields *x* and *y* are bound to a common object. The contours falling in *y* and *z* are not bound despite their collinearity and shared color. The solution to this binding problem requires sophisticated computation of surface order and border ownership and possibly knowledge about the nature of objects.

(B) The gray cross is moving coherently up and to the right (solid arrow). Motion sensors whose receptive fields lie at *x* and *y* cannot determine that the local displacement of contours is up–right. They process a component of motion that is either up or right (open arrows). The local contour motion is ambiguous until these features are bound to a common object. A partial solution to this problem has been shown to occur in the extrastriate visual cortex (area MT or V5), where feature motion detectors converge upon neurons with larger receptive fields (after Movshon et al., 1985).

(C) The interpretation of brightness at *x*, *y*, and *z* is affected through binding of *x* and *y* to a common plane. Although *x* and *z* are physically identical in their intensity, we see *x* as similar to *y*, because they represent surfaces of the same reflectance that are differently illuminated. The computation leading to this solution is believed to involve a comparision of intensities at corner junctions (after Adelson, 1993, 1999).

(D) Binding of different feature types, such as color and orientation, poses a problem for the nervous system to the extent that sensors

elaboration of this sample case by Eckhorn et al. (1988), and by Singer and his colleagues (Gray et al., 1989; Singer, 1993, 1994, 1999a; Singer and Gray, 1995; Engel et al., 1997; Roelfsema and Singer, 1998; Herculano-Houzel et al., 1999), into the "temporal binding hypothesis." These authors extended von der Malsburg's ideas by proposing that the solution of the binding problem is that populations of cells that represent low-level features (for example, neurons in primary visual cortex) synchronize their activity when they respond to different elements that are to be linked in the analysis of a scene. Singer's laboratory and others have presented supporting evidence, discussed in more detail below, that synchronous activity can be observed in these cells and that it can be modulated by stimulus features related to binding.

The Structure of the Binding Problem

There seem to be a number of problems with the idea of binding by a temporal code. Perhaps the most obvious is that the hypothesis is not a theory about how binding is *computed*; it is a theory only of how binding is *signaled*. The theory proposes that the result of the binding computation is represented by synchronous neuronal activity. This begs the question of what algorithms are actually used to solve the binding problem—how does the visual system decide which elements are part of single objects and which belong to different objects? To segment a complex image is computationally quite challenging, as is evident from the substantial literature on the topic in computer vision (e.g., Pal and Pal, 1993).

Image segmentation requires information from multiple sources. Simple features such as collinearity, color, texture, and common motion, which might easily be extracted from representations in primary visual cortex, only begin to define a solution to the problem, and even these feature-based combination rules must operate over wider areas of visual space than can plausibly be processed by neurons in V1. Consider the arrow segmentation cartoon of Figure 1A. Here, the cues of collinearity and common color seen by small V1 receptive fields at locations *y* and *z* would lead to an incorrect binding of contours to objects. Strategies for grouping in complex cases like the one shown must also take account of more highly elaborated representations that are not available in V1, representations that make explicit such "mid-level" features as the computation of pattern motion from more elementary motion information, as in Figure 1B (Movshon et al., 1985), or surface computations based on specific characteristic types of contour combinations, as in Figure 1C (Adelson, 1993, 1999). Full image segmentation probably requires even higher-level analyses, including the explicit inclusion of information from memory about the nature and structure of previously viewed objects and scenes.

It is important to distinguish binding object representations across space from the problem of binding particular feature qualities (e.g., color and orientation) to a

---

are not selective for conjunctions of features. If the orientation detectors do not signal color, how does the brain associate red with horizontal and green with vertical? This problem is thought to be solved by storing spatial tags, via attention (see Treisman, 1998), but a more complicated solution may be required for terminators, as in *z*, especially as the gratings are brought into proximity.

common object at a single spatial location. This latter form of binding is commonly discussed in the context of visual search for feature conjunctions (Treisman, 1996, 1998), and is likely to have a relatively straightforward solution because features are naturally assigned spatial tags in topographic representations within the visual cortex. Thus, while this form of the problem is not solved—we do not know how the brain recognizes conjunctions—it does not pose the same sort of combinatorial difficulties as binding across space. For example, there is no combinatorial explosion if a neural representation can encode whether there is a horizontal or vertical contour, or a red or a green one, at a particular location (Figure 1D).

Anatomical Considerations

It seems that the object binding problem cannot be solved in primary visual cortex, and that the computations involved cannot be completed until a fairly high level in the visual cortical hierarchy. The neurological literature supports the idea that binding is a high-level process. Visual binding deficits are vividly seen in Balint's syndrome, "an acquired disturbance of the ability to perceive the visual field as a whole, resulting in the unpredictable perception and recognition of only parts of it (simultagnosia)" (Damasio, 1985; Friedman-Hill et al., 1995; see also Rafal, 1997). Balint's syndrome is strongly linked to bilateral damage to the occipitoparietal region, including regions of the posterior parietal cortex associated with the direction of visual attention. Lesions to lower-order visual representations produce a variety of visual and perceptual deficits, but none seem obviously related to binding per se (Damasio, 1985).

If binding is not computed in the primary visual cortex, why is synchrony to be expected there? It might be that synchrony is imposed by feedback connections from the higher cortical areas in which the computation is done, but it is unclear what the utility would be of feeding back information of this kind. If the computations are done at higher levels of the system, then the answer is already represented there and does not need to be rerepresented at lower levels of the system. One reason could be that high-level representations are too general and too categorical to allow the tagging of individual stimulus features. For example, a neuron in temporal cortex that responds specifically to a particular object might not adequately identify all the contours and elements of that object. If such a neuron could control precisely organized feedback signals, these might tag the responses of primary cortex cells whose signals describe elements belonging to the object. But this kind of feedback would require exquisitely precise and dynamically configurable connections, not only to topographically specific parts of the primary visual cortex, but to particular orientation columns and laminae. The available data on feedback projections suggest that their organization is far too crude for this purpose (see Salin and Bullier, 1995, for a review). These projections seem designed to modulate neuronal activity in quite large regions of primary cortex, perhaps as a nonspecific "attentional spotlight" (Crick, 1984). There is no obvious way that such a coarse spotlight could be the source of a specific object-based synchronization signal that recruited specific neurons within the spotlight and excluded others; we see no evidence for the existence of the specific mechanisms needed to generate such feedback-based synchrony.

Binding by Synchrony and "Cardinal Neurons"

Another flaw in the temporal binding hypothesis concerns the question of which neurons would read the binding signal, and how. Singer, von der Malsburg, and others have emphasized the point that synchronized signals would be particularly effective in activating postsynaptic neurons that operate as coincidence detectors. We will discuss below our doubts about the utility of coincidence detection models for cortical neurons, but suppose for a moment that we grant the existence of these detectors. If there were special detectors configured to detect the coincident activity of particular groups of their input neurons, it is not clear how these would differ in any important way from the "cardinal" neurons whose implausibility led to the creation of the temporal binding hypothesis in the first place. Arguments can be made that the binding tag is not read by cardinal neurons but is instead used to modulate the activity of populations of neurons. But then the utility of synchrony is again unclear, because the population of output neurons (which modulate their firing rate in response to synchronous input) would lack the tags needed to represent which signals should and should not be combined for further processing. Synchrony is supposed to be special because it preferentially activates neurons that "read" the timing code. Usually, coincidence detectors are held to modulate their activity in response to synchronous inputs—this is how they are supposed to combine the conventional rate-modulated discharge to accord with their grouping. In any meaningful sense, this amounts to no more than a special way to create cardinal neurons.

Binding Is an Operational Signal

It is perhaps valuable at this point to restate the binding problem in operational terms. How are the signals from distinct populations of neurons combined for higher-level computation? How do high-level neurons determine which inputs carry information requiring further analysis? The problem would not arise if higher-order neurons received input from just a few selected sources, but the complexity of our sensory world and its neural representation seems to preclude this. It is more likely that higher-order neurons receive input from a variety of sources, representing, for example, many locations in space and many different local features. They must therefore determine which of these inputs carry signals worthy of further computation. The question that remains unexplained by the temporal binding hypothesis is how synchrony could usefully contribute to this determination.

Biological Plausibility of a Synchrony Code

Let us explore some implications of the proposal that synchronous spikes bind neural signals for subsequent operations. For this scheme to be useful, postsynaptic neurons must be able to distinguish synchronous from asynchronous spikes. Synchronous spikes stand out as special only if they do not arise frequently by chance. A receiving neuron cannot distinguish chance coincidences among asynchronous inputs from "special" synchronous spikes that would convey the additional information required for binding. How often would synchronous events occur among the inputs to a neuron

by chance? A general answer to this question is difficult, depending as it does on the number of inputs, their spike rates, and the time window in which synchrony is supposed to carry special meaning. A brief specific example is nonetheless instructive.

Synchrony in a Sparse Network

Consider a neuron that receives convergent input from an ensemble of ten neurons, each discharging asynchronously at an average rate of 10 spikes/s. By chance, any 1 input spike will occur within 5 ms of 3 other input spikes with an expected probability of 0.014. This is calculated as follows. The ensemble input rate is 100 spikes/s. From the time of 1 spike, the waiting time for 3 spikes is approximated by a gamma distribution. The waiting time is less than 5 ms for about 1.4% of the input spikes. Therefore, a neuron that emits a spike when at least 4 inputs arrive within a 5 ms window would discharge only 1 spike/s by chance. This is a sufficiently low background rate to allow us to identify spikes from this neuron as an indication that something special has occurred, leading to at least 4 spikes from the 10 input neurons within a short time interval. A spike from this hypothetical neuron would indicate that at least 4 of the 10 input neurons were active together. We cannot tell which group of 4, but the 210 possibilities (10 choose 4) could be further refined by combination with other neurons that received input from partially overlapping inputs. Coincidences of 3 events occur with a probability of nearly 10% and would occur about 10 times/s by chance. Coincidences among 5 or more inputs would very rarely occur by chance, but at this point the idea begins to lose relevance to the binding problem—by inventing a neuron that responds when more than half of its inputs are active, we have effectively solved the binding problem by connecting the right neurons and building a "cardinal" neuron.

We can still grant that if a neuron were to receive sparse excitatory input—meaning that very few excitatory input events arrive in the intervals between its own spikes—then it is reasonable to suppose that a set of excitatory inputs arriving within a short time of one another (say, within 5 ms) could lead to production of a spike sooner rather than later. This notion of coincidence detection was formulated clearly by Abeles (1982a). The idea seems only to be of value when cortical neurons operate in a regime characterized by low firing rates and sparse effective input. But when cortical cells operate in this regime, there *is* no binding problem—if firing rates are low and the number of effective inputs to a neuron is small, there is no need to tag them, or select among them, or multiplex a grouping signal with signals representing other attributes. We bring up these numbers to illustrate that the idea of spike synchrony as a signal *can* work, but only in a regime that seems inconsistent with what we know about the cerebral cortex.

Synchrony in Realistic Cortical Networks

In what sort of computational environment does a cortical neuron operate? This question may seem intractable in the face of computations that we do not fully comprehend, but the facts of cortical architecture provide guidance. Neurons in cortex receive 3,000–10,000 synaptic inputs, 85% of which are excitatory. Nearly half of the excitatory input to any one neuron comes from nearby neurons that fall within a 100–200 μm radius cylinder, arranged as a column, sometimes termed a mini-column (Szentagothai, 1973; Mountcastle, 1978; Braitenberg and Schuz, 1991; Peters and Sethares, 1991; Peters and Yilmaz, 1993). The remaining half can be traced mainly to a local network of horizontal connections originating in nearby cortex, leaving a minority of inputs from more remote cortical and subcortical structures. Within the 100–200 μm radius cylinder around a neuron there are roughly 1000 neurons that provide direct excitation. This number comes from an estimate of neuron density ($\sim 10^5$/mm^3) and from the observation that, on average, each neuron within the cylinder makes a contact with another neuron in the cylinder with probability of 0.09 (Braitenberg and Schuz, 1991; Hellwig et al., 1994).

These numbers suggest that cortical neurons receive abundant excitatory input and are embedded in a network of highly convergent signals. Because of the recurrent nature of the network, it is likely that most excitatory neurons receive similar inputs and emit spikes under similar conditions. This does not mean that all neurons in a mini-column fire identically, but only that the conditions that lead to a response of any one neuron are likely to involve considerable activity from a large number of its inputs and its targets. A conservative estimate is that each neuron receives several hundred excitatory input spikes for each spike it emits (Shadlen and Newsome, 1994, 1998). This number depends on a number of factors, such as cortical area and layer. The precise interpretation depends on factors influencing the efficacy of synaptic input, such as synaptic failure rates, synaptic adaptation, and dendritic amplification. But it nonetheless seems certain that cortical neurons receive many input spikes in the interval between one output spike and the next.

It therefore seems inevitable that many input spikes will arrive—in apparent synchrony—within any brief time window while the cortical column is active. Realizing this, what special significance can be attached to "synchronous" spikes? In a window of, say, 5–10 ms, the answer would appear to be none: in effect, *all spikes occur in synchrony with other spikes*. We cannot imagine how a neuron could sort the synchronous "binding" spikes from those that occur by virtue of cortical design.

Perhaps we should consider a shorter time window. The same numerical argument suggests that synchronous events are less common if we define a narrower window of synchrony. Perhaps a neuron could receive as few as 10–20 spikes in one tenth of an interspike interval. Synchronous spikes occurring within a millisecond or less could be regarded as distinctive, making a synchrony code on this time scale more plausible computationally. There are two problems with this idea. First, there is no biophysical evidence that cortical neurons can respond selectively to synchronous input of this precision. Unlike specialized neurons in the auditory brainstem, cortical neurons probably lack the biophysical mechanisms needed for precise coincidence detection at a millisecond time scale (Reyes and Fetz, 1993; Reyes et al., 1994, 1996; see also Koch, 1999, for a review). Second, reports of cortical activity with synchrony this precise are rare; experimental reports describe "synchronous" spikes in a somewhat broader window of 10–20 ms (e.g., Gray et al., 1989); in some cases, correlogram peaks 50 ms wide or more have

been taken as evidence for synchrony (e.g., Brecht et al., 1998). It is worth noting that in his original paper, von der Malsburg (1981), perhaps cognizant of some of the problems we have explored, suggested that the synchronous spikes carrying messages related to binding would need to occur within a much smaller window of about 3 ms.

Synchrony from Visual Inputs

Let us imagine that some mechanism, presently unknown, does exist that is capable of detecting synchronous activity in a window of 10–20 ms. The natural question that then arises is, does synchronization at this time scale uniquely signal binding? Empirically, this is equivalent to asking whether other processes in the nervous system produce activity that is synchronized with this same degree of precision. The answer to this question seems unequivocally to be "yes"—transient visual inputs are capable of eliciting bursts of cortical activity whose onset is reliable to within a few milliseconds, and whose duration can be as brief as 10 ms (Shadlen and Newsome, 1994; Bair and Koch, 1996; Bair et al., 1997; Buračas et al., 1998; see also Bair, 1999, for a review). The afferent volley of neural activity that follows brief visual events thus has a precision of synchrony that is at least as good as that reported for coherently oscillating groups of cortical neurons (e.g., Gray et al., 1989). So even if this degree of synchrony could be detected, there is no obvious way to discriminate visually elicited synchrony from synchrony due to binding.

Given the precision of visually elicited timing, it is natural to wonder whether the timing of visual events on a millisecond scale can influence perception. There are many examples of perceptual phenomena that depend on such timing. Our capacity to judge the relative position of two moving objects depends on sensory representation of both the place and time of the objects (e.g., see Burr, 1979; Fahle and Poggio, 1981; Carney et al., 1995). In a well-known example attributed to Pulfrich, a fine timing difference induced between the movement of features on the two retinae produces an apparent spatial disparity between the eyes, thus producing the sensation of stereoscopic depth (Carney et al., 1989; Howard and Rogers, 1995).

In the context of the present discussion, a number of groups have demonstrated that figure–ground segregation can be driven solely by temporal cues (Leonards et al., 1996; Alais et al., 1998; Usher and Donnelly, 1998; Lee and Blake, 1999). These experiments have shown, using a variety of displays, that the visual system can group features on the basis of similarity of time course, a natural extension of the Gestalt concept of "common fate"—a common modulation in time is an obvious cue for similarity (Köhler, 1930). It is reasonable to suppose that temporally coherent patterns of response would result from this common modulation, based on the simple premise that the time domain in the visual system is used to code the time course of visual events.

These demonstrations support von der Malsburg's original (1981) conjecture that temporal synchrony could be created by visual input. But the findings are difficult to reconcile with later versions of the temporal binding hypothesis, since these are based on the idea that synchrony binding would be produced by modulating an intrinsic oscillatory signal that is not time-locked to the visual stimulus. How is the brain supposed to distinguish the temporal modulation due to visual input from the temporal modulation produced intrinsically?

Kiper et al. (1996) explored this apparent conflict directly. They created displays in which groups of lines were segmented into regions by an orientation cue, and attempted to influence observers' segmentation performance by altering the temporal relationships among different elements of the display. Kiper et al. were unable to discern any influence of these temporal relationships on perceptual binding, suggesting that visual timing signals are processed independently of the spatial cues that give rise to binding; a similar conclusion was reached by Fahle and Koch (1995).

Thus, temporally precise visual activity is *sufficient* for binding, but it is not *necessary* for binding and its disruption does not affect binding elicited by other cues. Timing can indicate a common designation, but, like color and orientation, it can also be ignored (cf. Figure 1A). The natural conclusion is that whatever code the visual system uses for binding, that code does not rely on intrinsically synchronized neural activity.

Experimental Evidence on Binding by Synchrony in Visual Cortex

Characterizing Correlated Neural Activity

The main experimental evidence that supports the temporal binding hypothesis comes from studies of synchronous activity in recordings from spatially separated neurons in the visual cortex. Analysis in the main uses the cross-correlation technique (Perkel et al., 1967; Moore et al., 1970; Palm et al., 1988; Aertsen et al., 1989; Fetz et al., 1991), which reveals the probability of encountering a spike in one train as a function of the time lag before or after a spike in another train. Spike synchrony reveals itself as a peak in the cross-correlation between the spike trains near time lag 0.

Neurons can covary their discharge for many reasons. The simplest of these is spike timing covariation, that is, a tendency unrelated to firing rate for the two neurons to discharge at the same time—this corresponds to the common understanding and to our use of the term "synchrony." But factors other than synchrony can and do create peaks in cross-correlograms and many of these certainly contribute to many reported cases of correlated firing.

To demonstrate that correlogram peaks are not due to other kinds of spike rate modulation, the correlogram is typically corrected to remove the effects of that modulation. These "shuffle correction" techniques can only remove components from correlograms that are due to events that cause the same spike rate modulation on every repetition of an experiment, time-locked to some measurable event like the onset of a visual stimulus (Perkel et al., 1967; Melssen and Epping, 1987; Aertsen et al., 1989; Vaadia et al., 1991; Das and Gilbert, 1995). Should a pair of neurons change their discharge rate together, for whatever reason, in a manner that is *not* so time-locked, the cross-correlation will appear to show synchrony even after shuffle correction. Modulation of this kind appears in paired intracellular recordings from neurons in the visual cortex (Lampl et al., 1999). It is

not time-locked to experimental events and therefore cannot be removed by shuffle correction. Yet, it manifests itself in the cross-correlation and in all probability represents nothing more than shared variability due to the fact that inputs are themselves weakly correlated. This correlation prevents spontaneous fluctuations in afferent activity from being completely averaged out by synaptic convergence (Zohary et al., 1994; Mazurek and Shadlen, 1998, Soc. Neurosci., abstract; Shadlen and Newsome, 1998).

Experiments on visual cortical neurons almost always involve correlating visually driven activity. In this case, as Brody has recently shown (Brody, 1999a, 1999c), several factors other than spike timing covariation can produce peaks in cross-correlograms that give the appearance of synchrony. These factors include correlated slow response variability (expected if the overall excitability of the neurons varies from trial to trial in a correlated manner), correlated response latency (expected with moving stimuli if the position of the eyes with respect to the stimulus varies from trial to trial), and correlated visual responses associated with fixational eye movements (Bair and O'Keefe, 1998). All of these latter phenomena can be quite prevalent in visual cortical recording experiments; their contribution to the reported incidence of "synchrony" is difficult to establish but may be considerable (see Brody, 1998, for a striking example).

Determinants of Correlated Activity

Several groups have applied the cross-correlation technique to visual cortex to uncover functional interactions between groups of neurons. Notable among these studies are those of Toyama et al. (1981a, 1981b) and Ts'o et al. (1986), which showed that significant correlations between neurons are quite widespread in visual cortex and are strongest between neurons with related visual stimulus preferences. Correlated activity in cortical neurons is thus common and often serves as a functional indicator of shared connectivity.

The crucial question for the temporal binding hypothesis is whether correlated activity represents anything more than shared connectivity. The largest and most frequently cited body of evidence bearing on this question is the work of Singer and his colleagues. This group has published numerous studies of neural activity in the cortical and subcortical visual pathways, documenting extensive correlated activity both within and between visual areas. Gray et al. (1989; see also Engel et al., 1990) recorded from the striate cortex of anesthetized cats, and compared the degree of correlation between pairs of multiunit recording sites as a function of the separation between the recording sites and of the visual stimuli used. Their most striking finding was that the correlation between activity at moderately separated sites (about 7 mm) was strongest when a single stimulus covered both sites' receptive fields, weaker when similar but separated stimuli activated each receptive field, and weakest when independent stimuli moving in opposite directions activated each receptive field. This result is perhaps the single finding that propelled the temporal binding hypothesis into the foreground and is worth examining in more detail.

First, this particular observation was made in only a small number of cases out of a very much larger number studied—two cases are reported in Gray et al. (1989).

Engel et al. (1990) and Livingstone (1996) report a number of cases of correlation when the stimulating orientation is matched on the two receptive fields, but do not state whether any of their cases involve collinear stimulation with a single target. Second, the enhancement of correlation during stimulation by a long connected bar is the result that would have been predicted by the earlier results of Ts'o et al. (1986), who showed a decrease in correlation with distance—presumably, the use of a long bar "filled in" the gap between the sites with neurons that enhanced the correlated activity of the two remote sites (see also Das and Gilbert, 1995). Third, the reduced correlation observed when the bars moved in opposite directions would be expected if the time course of the responses was less oscillatory and if the receptive fields were less precisely aligned with the trajectory of the moving bars—differences in response time course and temporal structure can be seen in the published data (Gray et al., 1989, their Figure 2). Finally, the finding is of uncertain relevance to perceptual binding—many objects in the visual world give rise to images of separated contours that differ in orientation or motion (e.g., Figures 1A and 1B), yet under these conditions Gray et al. (1989; Engel et al., 1990) report that correlation is reduced or absent.

Correlation and Oscillatory Activity

The issue of synchrony is intimately bound up with the phenomenon of oscillatory firing in cortical neurons. Gray and Singer (1989) reported that at a substantial number of cortical recording sites, neurons and multiunit activity responded to visual stimuli by emitting clusters of spikes at intervals of about 25 ms, leading to a more or less regular rhythmic firing pattern at around 40 Hz. The prevalence of these oscillatory responses varies widely from laboratory to laboratory, for reasons that remain unresolved. While Singer's group and some others (e.g., Eckhorn et al., 1988; Livingstone, 1996) find oscillatory responses in about half their recordings, most others find their prevalence to be far lower, in the range of 2%–5% of recording sites (Tovee and Rolls, 1992a; Young et al., 1992; Bair et al., 1994; J. A. M., unpublished data). The reasons for this striking difference are unclear. Gray and McCormick (1996) reported that about 10%–15% of intracellularly recorded neurons gave oscillatory "chattering" firing in response to depolarizing currents; their chattering cells appear to be confined to cortical layer III. While these results and some earlier reports (e.g., Gray et al., 1989) suggest that oscillatory firing is of cortical origin, others have suggested an important role for oscillatory patterns relayed from the thalamus (Ghose and Freeman, 1992, 1997; Castelo-Branco et al., 1998).

The importance of oscillatory responses to interneuronal correlations was documented by König et al. (1995), who showed that while correlations between nearby neurons could be observed when the neurons emitted unpatterned responses, correlations between widely separated sites were very rarely observed in the absence of oscillatory responses. The low prevalence of oscillatory responses in most laboratories suggests that the dependence of these "long-range" correlations (i.e., correlations between neurons separated by more than 2 mm or so) on oscillatory responses may be a problem for the temporal binding hypothesis—if oscillations are only an occasional phenomenon, long-range correlations are likely to be rare and it is difficult to see how
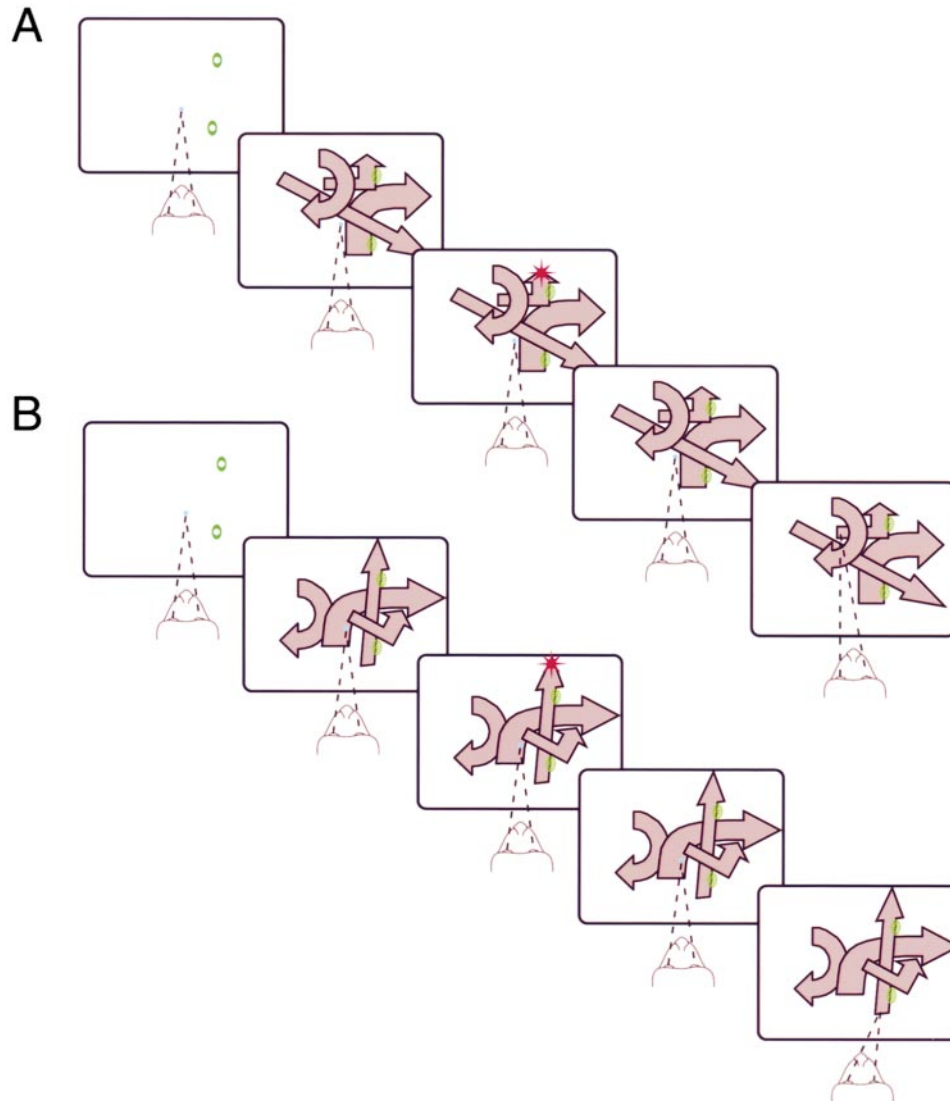
Figure 2. Experimental Task to Test the Temporal Correlation Hypothesis

A monkey is trained to fixate on a small spot (blue). Several overlapping arrows appear, one of which is cued transiently (red). After a delay period, the fixation point is extinguished and the monkey is required to shift gaze to the other end of the cued arrow. The arrangement of the objects is adjusted to stimulate two neurons in an early visual area whose receptive fields are shown by the broken ellipses in the task panels. (A) Features in the receptive fields belong to different objects. On correct trials (illustrated), the monkey indicates that the features are not part of the same arrow.

(B) Features in the receptive fields belong to the same arrow, as indicated by the monkey's gaze shift. The synchrony binding theory predicts that synchronous spikes, estimated by cross-correlation of the spike trains, will be more common in (B). The task is designed to allow the monkey to indicate that features are bound to particular objects. The monkey is rewarded for correct choices, but responses on error trials on such a task could be analyzed to see if changes in correlation predict binding errors. Other useful variants of this experiment are described in the text.

they could be a critical component of so important a perceptual process.

If, on the other hand, oscillations are ubiquitous, a different problem arises—the combination of oscillation frequencies and temporal precisions found in the literature would be capable of creating only a small number of independent assemblies. There is only 25 ms in one cycle of a 40 Hz oscillation. If "synchrony" is defined with a precision of 5–10 ms, this only allows time for a small handful of distinct epochs to be defined by different phase relationships. The informational gain to be obtained in this way seems rather meager.

Relating Synchrony to Perception

Two criticisms were quickly raised about the initial observations of Singer and his colleagues. First, the data were obtained from anesthetized animals and some were concerned that the correlations might be a consequence of anesthesia. Kreiter and Singer (1996) addressed this concern by documenting correlated responses by neurons in the middle temporal area (MT) of awake, behaving macaque monkeys, and by showing that these correlations could be modulated by stimulus features. Kreiter and Singer did not describe the kind of long-range correlations previously described in cat

V1; instead, they showed that neural activity in MT neurons with overlapping receptive fields was strongly correlated when a single bar activated both recording sites, and less correlated when the same sites were stimulated with two bars of different orientations. This result curiously contrasts with the reports of Movshon et al. (1985) and Rodman and Albright (1989), who showed that many MT neurons combine signals from multiple orientations to compute "pattern motion." It seems inconsistent with the temporal binding hypothesis that this form of low-level perceptual binding (schematized in Figure 1B) is associated not with an increase but with a *decrease* in correlated firing.

The second criticism is that because no perceptual judgments were made during the experiments, evidence that the chosen stimulus configurations actually promoted perceptual binding was circumstantial—the experiments typically used stimuli that promoted binding-like effects in human observers, but did not establish that experimental animals perceived the stimuli in the same way. This concern has not been directly addressed by Singer and his colleagues, but they have conducted several studies in an effort to approach the issue of perceptual relevance.

Roelfsema et al. (1994) studied response correlations in the visual cortex of esotropic cats (i.e., cats with convergent strabismus) which developed amblyopia, a developmentally induced disorder of vision thought to reflect changes in visual cortical function (reviewed by Kiorpes and Movshon, 1996; Kiorpes and McKee, 1999). They reported that response correlations between neurons driven by the amblyopic eye were weaker than between neurons driven by the fellow eye, and concluded that the loss of neuronal synchronization was the substrate of the amblyopia. The connection between amblyopia and temporal binding is potentially of interest, since some aspects of the abnormalities of human amblyopic vision suggest a deficit of perceptual binding (Hess et al., 1978; Polat et al., 1997; but see also Levi and Sharma, 1998).

The study of Roelfsema and his colleagues is therefore topical but is weakened by a number of deficiencies. Since esotropia produces amblyopia only in a fraction of subjects, amblyopia needs to be shown for each animal. The documentation of amblyopia in this study was incomplete, and the main behavioral finding was of rather profound visual deficits in *both* eyes. Moreover, although they do not actually show the data, Roelfsema and his colleagues claimed to find no difference in the response properties or spatial resolution of cells representing the two eyes; reports on amblyopia in both cats and monkeys suggest that such changes are a frequent and perhaps invariable component of amblyopia (Eggers and Blakemore, 1978; Chino et al., 1983, 1991; Movshon et al., 1987; Crewther and Crewther, 1990; Kiorpes et al., 1998). It is possible that Roelfsema and colleagues' animals were not true unilateral amblyopes, suggesting that the reported effects on interneuronal correlation may have been related to factors other than amblyopia. It is also notable that the difference in correlation between the eyes reached significance in only two of four animals tested with bar stimuli, and in only one of three animals tested with grating stimuli. Therefore, the results, while suggestive, cannot be taken to establish a secure link between perception and correlated neuronal activity.

Fries et al. (1997) used a different method to approach the question of perceptual relevance. Using cats whose cortical binocularity was reduced by strabismus, they studied cortical responses and correlated activity while the two eyes viewed different stimuli. These conditions provoke binocular rivalry in human observers—the two eyes' views are not seen together; rather, one or the other dominates perception. In separate experiments using an eye movement technique on the same animals, Fries and his colleagues established conditions in which one eye's percept was reliably dominant. They showed that under conditions in which this eye was presumed to be dominant, cortical correlations were enhanced. When the tested eye was presumed to be suppressed, correlation was reduced. The authors suggest that this result reveals a role for correlated activity in perceptual dominance during binocular rivalry. Regrettably, the neural measurements were not made at the same time as the rivalry measurements, nor were the visual stimulation conditions similar for the two cases. Thus, the claim that the changes in correlation were due to perceptual changes can only be taken as an intriguing conjecture and not as a proven fact.

Perceptual States Are Signaled by Changes in Firing Rate, Not Correlation

One of the implications of the temporal binding hypothesis is that changes in perceptual state associated with binding phenomena might not be revealed by a conventional analysis of neuronal firing rate. The reasoning is that firing rate provides the afferent perceptual data, while synchrony provides the code that glues those data together into percepts (e.g., deCharms and Merzenich, 1996; Fries et al., 1997). There are, however, many demonstrations that a rate code carries behaviorally relevant information when animals perform perceptual tasks related to binding, and a few demonstrations that synchrony carries no information under the same conditions.

Several groups have studied neuronal activity in behaving animals making perceptual judgments. Britten et al. (1992, 1996; Shadlen et al., 1996) studied the perception of global motion in random dot kinematograms. While this is not a traditional binding task, the perceptual judgment in these studies requires integration of common local motion signals over space and is arguably a form of binding. Britten and his colleagues showed that stimulus-induced modulations of firing rate could account for animals' psychophysical performance. More pertinently, they showed that when the visual stimulus conditions were unvarying, changes in psychophysical performance were correlated with changes in firing rate (Britten et al., 1996). Bradley et al. (1998) studied a structure-from-motion discrimination task and again showed that in the absence of unambiguous cues to depth, the animals' perceptual judgments were reflected in changes in firing rate. Logothetis and his colleagues have studied neural activity while animals reported which eye controlled perception during binocular rivalry (Logothetis and Schall, 1989a; Leopold and Logothetis, 1996; Sheinberg and Logothetis, 1997; Logothetis, 1998). They found that neurons in several visual areas modulated their firing rates in association with changes in

rivalry dominance, which occur without changes in the visual stimulation conditions. These results are more compelling than the rivalry experiments of Fries et al. (1997, discussed above), because Logothetis and his colleagues directly compared neural activity and behavioral state on every trial.

Zohary et al. (1994) used the paradigm of Britten et al. (1992) to study interneuronal correlations during psychophysical performance. Analysis of their data reveals that variations in psychophysical performance are not associated with changes in the synchrony of firing between neurons; moreover, the more salient—and arguably more strongly bound—high-coherence stimuli in their studies elicited slightly weaker synchrony than the low-coherence ones (W. Bair, E. Zohary, and W. T. Newsome, personal communication).

Lamme and Spekreijse (1998) offered what is perhaps the most direct test of the temporal binding hypothesis to date. They recorded from separated groups of neurons in macaque visual cortex while monkeys viewed (and made judgments about) a figure–ground display created from oriented line segments. While keeping the receptive field stimulation constant, they measured correlated activity when the two recording sites represented the same region (figure or ground) and contrasted that with correlated activity when the two recording sites represented different regions of the display; they found no differences in correlated activity between the two conditions. Lamme and Spekreijse did not attempt to relate correlated activity directly to perception by analyzing trials on which the animals made errors, and it is notable that the peaks in their published correlograms are so broad that they may not indicate synchrony in the sense of spike timing covariation, as discussed above (Brody, 1999c). Nonetheless, the clarity of their negative findings is striking.

In summary, the results of these studies all suggest that changes in perceptual state are not associated with changes in correlated cortical activity. At least when recordings are made from areas outside the primary visual cortex, perceptual changes are invariably associated with changes in firing rate. It is of course not necessary to the temporal binding hypothesis that rate modulations be absent when perceptual changes occur. Nonetheless, it is striking that perceptual changes *are* associated with changes in firing rate and *are not* associated with changes in response correlations.

Rigorously Testing the Temporal Binding Hypothesis

Most of the data that are usually held to support the temporal binding hypothesis prove to be flawed or circumstantial; experiments that seem to test the hypothesis more directly generally yield negative results. Still, it can be argued that, despite more than 10 years of research, the right experiment has not been done. What kind of experiment might we devise to provide a true test of the hypothesis?

In Figure 2, we outline a proposed experimental design, which has some features in common with that of Lamme and Spekreijse (1998). Recall the binding display caricatured in Figure 1A. This nest of superimposed arrows cannot be correctly perceived unless the separate parts of the objects are bound together. Operationally, we can learn whether binding occurs by asking the experimental animal to select with an eye movement the tail of an object whose head has been cued. Suppose we record from two neurons with separated receptive fields (green ellipses). We can align the display so that the two receptive fields either do or do not fall on contours belonging to the same object (Figures 2A and 2B) and then compare neuronal correlations between these two conditions.

This design would make it possible to explore a variety of factors that have been associated with binding. The experiment can be done with receptive fields of any location and orientation preference by suitably adjusting the configuration of the figures, to assess the role of collinearity and shared orientation preference. The role of contour connection and occlusion can be determined by using occluded or unoccluded objects to stimulate the two receptive fields. The role of focal attention and movement preparation in binding can be explored by comparing conditions in which the cued object falls within one or both receptive fields with conditions in which an uncued object stimulates the neurons. The validity of correlation change as a predictor of behavioral performance can be established by using suitable single-trial statistics to compare activity on "correct" and "error" trials under the same stimulus conditions (cf. Britten et al., 1996). The temporal binding hypothesis makes clear predictions for all these cases, and both positive and negative findings could be clearly discerned and interpreted.

Binding without Synchrony

In the preceding sections, we examined the temporal binding hypothesis and reevaluated many of the data that have been held to support it. We now turn to the issue that led to the hypothesis in the first place, the idea that binding is a special problem that cannot be solved by "conventional" mechanisms of neural signaling (von der Malsburg, 1981). Such "conventional" thinking would postulate that, like other complex problems in perception, the binding problem is solved by the successive elaboration of progressively more complex representations of visual scenes. It does not seem that there is any good reason to abandon this strategy. We need not embrace "grandmother cell" theories but simply consider that higher cortical areas encode scene attributes and object identity, along with other results of perceptual analysis, in the distributed rate-encoded activity of populations of neurons.

While explanations of this kind have not received extensive attention, at least one serious attempt has been made to explain feature binding in this way (Olshausen et al., 1993, 1995; Tsotsos, 1995). This theory postulates convergence to cells in higher cortical areas but acknowledges that some problems cannot easily be solved by simple convergence. These problems are addressed by special circuits, controlled by the pulvinar, which act dynamically to alter the flow and combination of visual signals. The "binding problem" in this theory is implicitly solved in higher cortical areas that receive the dynamically routed and recombined information. No specially coded form of neural activity is required.

Solving the Binding Problem in the Context of Action

We believe that the visual cortex may not be the only place to solve the binding problem; nor is it the appropriate place to represent the solution. A more sensible

locus is the association cortex, more specifically the posterior parietal cortex. We are guided by the clinical observation that failure of binding occurs with damage of the parietal cortex (Damasio, 1985; Friedman-Hill et al., 1995; Rafal, 1997; Robertson et al., 1997). While our understanding of the neurophysiology of posterior parietal cortex lags behind that of the visual cortex, what is known seems compatible with the idea that the parietal cortex computes a solution to the binding problem.

There is increasing evidence that the parietal cortex organizes information from the visual cortex in terms of its significance for behavior. For example, neurons in the lateral intraparietal area (LIP) respond to a variety of visual targets when they specify the location of an intended eye movement (Andersen, 1995; Colby et al., 1996; Platt and Glimcher, 1997). Neurons in more medial cortex respond before reaching movements to targets (Caminiti et al., 1996; Snyder et al., 1997), and, more anteriorly, Sakata and colleagues have described neurons that respond to shapes that elicit particular grasping postures (Sakata et al., 1995; Murata et al., 1996). In each of these cases, the neuron responds to visual stimuli in a manner that reflects their salience with regard to an action or behavior (Rizzolatti et al., 1997; Colby and Goldberg, 1999).

These observations are related to binding because the designation of salience can be based on information derived from a separate spatial location—a visual stimulus within the receptive field of a parietal neuron can be designated as salient by another stimulus outside the neuron's receptive field. This property is captured in experiments in which a monkey is instructed to make eye movements to one of two visual targets depending on properties of a third stimulus placed elsewhere in the visual field. For example, neurons in parietal cortex respond to a visual stimulus within their receptive fields and also to a complex motion stimulus placed *outside* the receptive field, which serves to instruct a gaze shift to the receptive field stimulus (Shadlen and Newsome, 1996). These neurons encode the binding between one sensory stimulus (the instruction) and a second, behaviorally relevant stimulus within the receptive field (Assad and Maunsell, 1995; Platt and Glimcher, 1997; Eskandar and Assad, 1999; Horwitz and Newsome, 1999; Kim and Shadlen, 1999; see also Leon and Shadlen, 1998, for a review).

How do such observations relate to the binding of object features? We do not imagine that binding is solved by making movements of the eyes and limbs. However, neural circuits that guide reaching and gaze depend on a solution to the binding problem. The control of grasp, for example, presupposes knowledge of spatial relationships between features and the object to which they are bound. Lifting a pencil by its eraser or a teacup by its handle, or shifting gaze to the tail of an arrow based on the sight of the tip (Figure 2), all require that the binding problem be solved. We propose that the requisite computations, which begin in primary visual cortex and are elaborated in successive visual representations, are completed and synthesized in the parietal lobe. Specifically, the parietal cortex determines the relevance of local features to particular behaviors, based on instructions from other locations in the visual field

and more generally from other sources. In this context, then, binding is the designation of spatial salience of a local feature in the neural receptive field on the basis of another feature or features located outside the receptive field. This notion of binding is specifically related to neural circuits that organize behavior, such as the target of a gaze shift or a reach movement. In this sense, binding can be viewed as a grouping of features that leads to the designation of behavioral targets.

This style of computation has the structure of a hypothesis test, strongly reminiscent of von Helmholtz's (1925) and Gregory's (1970) ideas on how perceptual information is grouped to generate the perception of objects and scenes. Parietal computation evaluates the evidence for a specific behavioral act that would be directed to the neuron's response field. For example, a working definition of the response field for a neuron in LIP is the locus of stimulus positions that cause the neuron to respond, especially when the stimuli comprise a target for a saccadic eye movement. The evaluation is based on perceptual data, learned associations, and predictive coding. An example of a learned association is a spatial cue that directs an eye movement to a particular target. In the experiments of Shadlen and Newsome, a monkey observes a visual display and learns to shift its gaze to a particular target based on the display's properties. Neurons in LIP indicate qualitatively by their response to the visual display whether or not the target in the response field is the one that will be chosen, and quantitatively the degree to which perceptual evidence supports this choice. This response pattern is perfectly formed to represent the generation and evaluation of a perceptual hypothesis (Shadlen and Newsome, 1996; Kim and Shadlen, 1999).

Our proposal represents only the skeleton of an alternative to the various theories we considered above. To put solid theoretical flesh on its bones will require substantial refinement and experimental scrutiny. For example, feature binding for the purpose of reaching and scanning may not be the same as binding for perceptual awareness—the neural circuits that mediate object identification may be distinct from those that control grasp and gaze (Ungerleider and Mishkin, 1982; Goodale, 1993; Goodale and Humphrey, 1998; but see also Franz et al., 1999). We have not specified with precision how the computation is performed—yet it is clear that the calculations involved can be done within the "conventional" framework of rate-modulated neural representations, without recourse to a new and special kind of signal. These ideas are also attractive because they are consistent with neurological evidence on parietal lobe damage. This formulation also tames the combinatorial explosion of features and objects because the groupings are constrained by possible action (Churchland et al., 1994). Features are bound together to the extent that one feature can be viewed as an instruction to act in some way upon another.

Conclusion

We have presented a critical evaluation of the hypothesis that a temporal code based on synchrony of spike timing represents the process of feature binding. We considered several lines of reasoning that suggest that

the theory is inadequate in conception and impoverished in support. The theory is incomplete in that it describes the *signature* of binding without detailing how binding is computed. Moreover, while the theory is proposed for early stages of cortical processing, both neurological evidence and the perceptual facts of binding suggest that it must be a high-level computation. Consideration of the architecture of the cerebral cortex suggest that it lacks the mechanisms needed to decode synchronous spikes and to treat them as a special code. There is ample experimental evidence for correlated cortical activity but little that directly or compellingly links this activity to binding. In contrast, there is considerable evidence that the rate-modulated activity of cortical cell populations is crucial in mediating perceptual binding.

The conclusion seems inescapable that the theory as proposed is untenable. Nonetheless, the theory has sparked renewed interest in the problem of binding and has provoked a great deal of important research. It has also highlighted the crucial question of neural timing and the role of time in nervous system function. The problems that gave rise to the theory are still important problems that remain to be solved, and it is certain that the efforts of the theory's proponents and opponents will advance our knowledge both of higher visual functions and of the algorithms used by that most enigmatic of computers, the cerebral cortex.

References

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