The What and Why of Binding: Review
The Modeler’s Perspective

Christoph von der Malsburg*
Institut für Neuroinformatik
Ruhr-Universität Bochum
44780 Bochum
Federal Republic of Germany
University of Southern California
Los Angeles, California 90089

In attempts to formulate a computational understanding of brain function, one of the fundamental concerns is the data structure by which the brain represents information. For many decades, a conceptual framework has dominated the thinking of both brain modelers and neurobiologists. That framework is referred to here as “classical neural networks.” It is well supported by experimental data, although it may be incomplete. A characterization of this framework will be offered in the next section.

Difficulties in modeling important functional aspects of the brain on the basis of classical neural networks alone have led to the recognition that another, general mechanism must be invoked to explain brain function. That mechanism I call “binding.” Binding by neural signal synchrony had been mentioned several times in the literature (Legéndy, 1970; Milner, 1974) before it was fully formulated as a general phenomenon (von der Malsburg, 1981). Although experimental evidence for neural synchrony was soon found, the idea was largely ignored for many years. Only recently has it become a topic of animated discussion. In what follows, I will summarize the nature and the roots of the idea of binding, especially of temporal binding, and will discuss some of the objections raised against it.

Classical Neural Networks
Classical neural networks were developed as models of brain function. In developing these models, several questions needed to be addressed:
(1) How are brain states to be interpreted as representations of actual situations? In other words, how is neural activity interpreted as a neural code, or, in computer parlance, as a data structure?
(2) What is the nature of the mechanisms by which brain states are organized?
(3) In what format is information laid down permanently in the brain?
(4) How is memory laid down? In other words, what are the mechanisms of learning?

The questions remain open, but, as we shall see, plausible answers have been offered by classical neural networks. Interestingly, in the context of the field of Artificial Intelligence, no general answers to these questions are provided, in the conviction that specific problems need specific data structures and specific algorithms (or, in our parlance, mechanisms of organization). Neuroscience and neural modeling, on the other hand, have the ambition to find general answers. It is this commitment to generality that results in the binding problem being a fundamental feature of the neural code.

The Psychological Moment
Before discussing the answers to the above questions that are postulated by classical neural networks, it is important to introduce an important parameter. Although not often made explicit, it is important to fix a temporal scale \( T \), which we will refer to as the “psychological moment” (in the sense of “short period”). At times shorter than \( T \), one speaks of mental state or brain state, whereas at times greater than \( T \) one sees a succession of states or a “state history.” Whereas state history is subject to conscious scrutiny (that is, it potentially gets reflected in all modalities—memory, language, etc.), no such conscious analysis is possible below \( T \). State history is ignored by most or all models, and the conceptual disagreement about the binding issue focuses exclusively on the definition of state, that is, on times below \( T \). It is difficult to pin down a definite value for \( T \), but a plausible range may be from 50 to 200 msec. Regardless of the exact value, it is important to realize that the arena for the discussion of the binding problem is at a time scale less than \( T \).

Classical Neural Networks: Answers to the Issues of Brain Function
Classical neural networks are described by a deeply engrained set of concepts, often attributed to Hebb (1949) and Hayek (1952) but in reality much older, which give definite answers to the questions 1–4 above.

(A) The neural code: neurons are taken as concrete symbols, as semantic atoms. They can be interpreted in relation to patterns and events external to the organism. Neurophysiology has provided a solid experimental basis for this statement, although some extrapolation is needed to extend it to all neurons in the brain. The interpretation of neurons as semantic atoms is generally accepted and is not the matter of much dispute. However, the following addition will have to be a focus of our discussion:

(A') A neuron has only one degree of freedom for a given psychological moment: it is either on or off (or it is on to a certain degree). Thus, the brain state is described by a list—or vector—of neural activities. In order to know what the brain is about in a given psychological moment, it is only necessary to know this vector, as well as, of course, a description of the symbolic meanings of all neurons. This interpretation of brain activity deliberately ignores the fact that actually recorded neural signals are not constant in any sense over \( T \)—or even over any fixed time scale, for that matter. It is rather maintained that the observed microstate history is inconsequential for the function of the brain.

(B) The mechanism of organization of brain states is based on the fluxes of excitation and inhibition, a neuron collecting incoming signals and firing when a threshold is surpassed (see, for instance, the switching rule of McCulloch and Pitts [1943]). The dynamics of the system...
If two sets of active neurons (left and middle panel) are simultaneously activated (right panel), information on their membership in the original sets is automatically lost.

Let’s assume it was vital for an organism to trigger some action in response to a triangle if it was in an upper position but not in a lower one. The reaction would then have to be tied to the coincidence of activity in cells (triangle) and (top), which, however, would also occur if the triangle were at the bottom and a square at the top. The animal therefore would respond to a so-called false conjunction, perhaps with grave consequences. An analogous situation occurs in the brain. Correspondence between object type and position is explicit on the retinal level. Its loss on the way to the output of the circuit is due to the generalization that is taking place within the circuit: for instance, in the brain’s “what” and “where” pathways in the temporal and parietal pathways of primate cortex.

This problem is a general one, with implications far beyond that of vision. Imagine a mental object that is represented by the set \( P \) of neurons (refer to Figure 2) and another mental object represented by set \( Q \) (possibly overlapping with \( P \), but that is not a point here). Now it becomes important to activate both objects in the same mental operation (when, for instance, comparing them). What would be more natural than to coactivate both sets in the same brain state? Such coactivation, however, leads to what we call the “superposition catastrophe”: the two sets will merge into one, and the neural code will not express the information needed to subdivide the composite state into its components (see Figure 2).

**Combination Coding as a Solution to the Binding Problem**

Rosenblatt’s problem has a simple solution in terms of combination-coding cells. It would suffice if there existed a neuron that reacted to a triangle in the top
position (or to a set \( P \) or \( Q \) of neurons). This could be realized with the help of connections from the lower, elemental levels, on which generalization has not yet taken place. However, a problem arises where appropriate combination-coding cells do not exist or cannot exist (due to the impracticality of the large numbers required, or to the previous history of the system) and where there are connection patterns that could cause confusion. The assumption that combination coding cells are available when and where required is problematic when the system to be modeled is a general purpose device.

Most symbol systems have means of combining elementary symbols into more complex ones, which can then be handled as units without danger of ambiguity and which have explicit structure on the basis of which they can be compared, recognized, decomposed, and further combined to build even higher structures. That the classical code of neural networks doesn’t have such means is the root of the binding problem. It is a very curious proposition that the brain, the ultimate handler of symbol systems, shouldn’t have a general mechanism for combining subsymbols. Other symbol systems (such as mathematics or natural languages) suggest that more complex binding patterns may be required than just grouping a number of elements into one block with no internal structure. The visual image of an extended object may need to be represented as an array of local features that are bound together in a way that expresses the topological neighborhood relationships within the figure or even a hierarchy of object parts (Biederman, 1987); similarly, the representation of natural language structures requires binding arrangements with hierarchical structure.

Why not have a purely hierarchical system of combination-coding cells? There may be good reasons that certain combination-coding cells should not exist. The more complex the combination a cell represents, the more special the context to which it refers. Any experience gained in a particular context (and recorded in terms of changed synaptic connections) should be affixed to the most general description of the situation, so it can be exploited in other contexts. If, for instance, something was associated with the appearance of a triangle that just happened to be in an upper position, it would be inefficient to affix the association only to an upper-triangle cell, for then the experience would have to be repeated and relaid for all positions of the triangle. In a similar vein, the absolute location of a speck of ink on paper is of very little relevance; what counts is its relative location in a pattern. Our environment is complex because it is combinatorial: complex objects and situations are constructed by combining simpler elements. To try to represent this complexity in a non-combinatorial way, letting single cells stand for external objects of any complexity, appears to be a terribly inefficient strategy. Any combinatorial symbol system, however, needs a mechanism to bind elements into groups. Some evidence exists that suggests that the brain does not simply use combination-coding cells to process stimuli. Psychophysical studies show that under certain conditions the brain’s binding mechanism may fail, and people will report “illusory conjunctions” (see Wolfe and Cave, 1999 [this issue of Neuron]). When given enough viewing time, subjects do not commit such “false conjunction” errors. In a related vein, conjunction search experiments (Treisman and Gelade, 1980), in which subjects are asked to find an object with a specific combination of features in an array of distractors with one of those features each, show that reaction times scale linearly with the total number of elements in the display. From these types of experiments, one can conclude that relevant combinations of features are not represented by combination-coding cells and that the brain requires time to form or ascertain the correct bindings.

Are Classical Neural Networks Universal? There is a widespread opinion that classical neural networks are a universal medium with no limits to their abilities and that consequently they are not subject to the binding problem. I will address this claim in two steps. First, I will discuss whether universality suffices as a solution to the brain’s problems and then I will raise doubts as to whether classical neural networks are indeed a universal medium.

The idea of universality was crystallized with Alan Turing’s formulation of the Turing machine and his demonstration that no effective procedure can be conceived that cannot be realized as the program of a Turing machine. Thus, any completely specified function can be realized as a program, or algorithm, run on a computer, the only limits to this being storage space and time. From this, it was extrapolated that mental processes, if only made concrete in terms of rules, could be realizable in machines. Under this view, the brain is a digital machine and has the same universality as the computer or the Turing machine, if only sufficiently many neurons are available. McCulloch and Pitts (1943) applied this idea to modeling of the nervous system, proving that any logical function can be realized as a network of threshold elements if they are appropriately connected.

But what does universality buy? It can be compared to the universality of a pen and sufficiently many sheets of white paper as a universal medium for formulating novels. You still have to write them. Over time, the field of Artificial Intelligence discovered that it is not a practical task at all to write a program that emulates the capabilities of the brain. It is becoming increasingly clear that the only goal we can hope for is to establish a system that constitutes a basis for self-organization and learning, as the equivalent of a newborn, or better still, the genetic program for the development of a brain, and to let it learn from experience and from communication with others.

Brain theorists realized this in the late 1950s and modified McCulloch and Pitts’ networks to accommodate self-organization and learning. The resulting framework of classical neural networks has a tendency to fall into stable patterns and learns by synaptic plasticity. However, these changes may have come at a price: it is not clear whether neural networks are universal in any sense, although the community seems to have inherited the implicit belief that they are and that any brain function can be modeled on the basis of those few abstractions from the real nervous system that went into the formulation of neural networks. It is not even clear how to formulate a new universality theorem. Classical universality states, “give me a procedure and I’ll tell you how to implement it.” The neural network version of universality would have to be, “give me a brain problem
and I will be able to implement it in classical neural networks. But how can we characterize brain problems in any general and satisfactory way? It would be foolish to argue that ‘this is a particular problem I have solved on the basis of classical neural networks, which proves that all of them can be solved this way.’

In the context of the present discussion, the particular version of universality that some critics of the binding problem uphold, is ‘state a concrete problem, and I will solve it with a classical neural network without ever running into a binding ambiguity.’ The common argument to support this posits that any binding ambiguity implicit in a stated problem can be dealt with by the concrete combination-coding cells in the model network that solves it (for example, see Riesenhuber and Poggio, 1999a [this issue of Neuron]). A concrete example of this approach is Mel and Fiser’s model (1999) for recognizing words from text. The network codes text in terms of triplets of contiguous letters and is based on the statistical observation that in English no two words agree in all their letter triplets. Thus, in this special case, the model completely avoids the binding problem and ensuring compositional ambiguity that would berid a model based on a representation by single-letter cells only.

Although such examples are meant to support a general universality claim, it is very doubtful that such a claim can ever be established. It is too easy to state problems that are far beyond the abilities of present neural network models. Just think of the task, ‘emulate the human ability to segment visual scenes, with all the necessary cue integration.’ Although the problem is old, there exists no classical neural network solution to it, and perhaps for good reason.

Binding by Temporal Synchrony

How could binding be implemented in the brain? The basic idea of temporal binding is that signals of neurons that are to be grouped together are correlated in time. Neural signals can thus be evaluated in two ways: one of them is the classical concept of neural firing rate, in which the relevant parameter is the running average of the number of spikes arriving within any period $T$. The second concerns temporal correlations of signal fluctuations happening on time scales faster than $T$, and it is these correlations that express binding. The subdivision of the time scales above and below $T$ is on final account arbitrary (unless one sticks to the distinction that changes slower than $T$ are accessible to introspection while faster changes are not). Pick a scale $T$, and then evaluate signal fluctuations below $T$ in terms of correlations, while calling fluctuations above $T$ rate changes (there is, of course, a lower bound, that depends on the fastest temporal scale that can be processed by neural tissue). For the brain, $T$ is related to the psychological moment, ill-defined as it may be, so we will take $T$ to be of the order of 50 to 200 msec (although scales up to minutes and beyond are also of potential relevance). Throughout this discussion, I discuss signal correlations as if they were to be evaluated without taking into account relative delays. However, it may be necessary to also consider delayed coincidences as argued in Bienstock (1995).

To understand the general idea of temporal binding, the exact nature of the signal fluctuations is not relevant. There has been much discussion over the experimental evidence concerning whether oscillatory signals are or are not important in this respect, but that discussion is a side issue that shouldn’t cloud the more fundamental question of whether binding is a problem and whether neural signal correlations are a solution to that problem. Although much of the discussion of signal correlations focuses on the binary case, the correlation of just two neurons, it should be emphasized that the much more relevant and important type of event concerns correlations of higher order: the simultaneous firing of larger groups of neurons. The reason is this: for correlations to play an important role in the brain it must be possible for them to be evaluated quickly and reliably. Since for any given set of impulses there may be accidental coincidences, it is vitally important to discern true correlations from noisy background. For binary correlations, this is only possible for long observation times, but for correlations of sufficiently high order, even individual coincident events can become highly significant.

With regard to the plausibility of temporal binding, there are the several fundamental questions that need to be discussed:

1. How do correlations in temporal signal structure arise? Ultimately, the purpose of temporal binding is to express significant relations between data items, e.g., of causal or spatial nature; the physical interactions establishing such relations must be represented by signal correlations. Since many of these interactions are already present in the external world, temporal correlations can be imposed by external stimuli triggering the neural signals. That this happens with causally related external events is evident, but it is less often realized that the same can be due to our own bodily and eye movements, which create a stream of sensory impressions whose temporal structure expresses spatial structure of the environment. Thus, some of the signal correlations relevant for binding are already implicit in the perceptual input (see also Singer, 1999b [this issue of Neuron]).

As a short aside, the argument is often raised that the use of temporal patterns for expressing binding may clash with the use of temporal patterns for other purposes, such as the representation of temporal structure as given in the external world. This is especially relevant to the auditory, language, and motor modalities. This clash may be avoided by the nervous system by recoding temporal signals into a format that does not involve rapid signal changes. Single neurons responding to and representing syllables would be an example of this.

More correlations (probably the overwhelming majority) are created within the nervous system by synaptic connections. If neuron a fired neuron b, the signals of the two would be correlated (disregarding a small delay). Correlations induced by synaptic connections also signify causality. In addition, activity in neurons without connections between them but with connections from a common input can be correlated. In the Rosenblatt example, the binding problem could be solved if the neurons in V1 that are activated by a triangle in a given position pass the temporal signature of their signals on to neurons expressing shape identity on the one hand
and position on the other, such that those signals came to express their common origin.

(2) How are signal correlations evaluated in the brain? If two action potentials arrive at a common target neuron, their relative timing exerts a strong influence. If signals arrive simultaneously, they can cooperate to raise the neuronal membrane potential above firing threshold. If, however, they miss each other in time, so that the effect of the first impulse has decayed before the second arrives, they might both fail to fire the target neuron. Thus, neurons act as coincidence detectors and do evaluate signal correlations (Abeles, 1982b; Marsalek et al., 1997). The exact details of this interaction depend upon many complex factors, including membrane time constants, nonlinear effects, and dendritic geometry. In consequence, current neurophysiology cannot solidly predict the temporal resolution at which spike coincidences are evaluated; however, a likely range is 1–10 msec.

If all correlations were to be evaluated globally by single neurons, a combination-coding cell would be required for each binding pattern, defeating the purpose of binding. However, complex correlation patterns created by a circuit of interconnected neurons can be evaluated by other circuits of appropriately interconnected neurons, each individual neuron checking only a small subpattern. Thus, pairs of circuits may or may not resonate with each other in terms of the correlation patterns that they produce. This point is probably most easily understood with reference to the concrete models of invariant object recognition that are discussed below.

It was proposed (von der Malsburg, 1981) that correlation patterns are also evaluated by rapid reversible synaptic plasticity (in addition to slow plasticity). A connection that is physically present and would cause confusion in a given situation could be temporarily inactivated when activity on both the presynaptic and postsynaptic side is sensed, but is uncorrelated. Confusion could thus be suppressed in the given situation, even if the signals involved were to develop stray coincidences, until the switched-off connection returned to near its previous value on the time scale of seconds or minutes.

(3) How can correlation patterns be effective on physiological timescales? If they are to play a role in the brain’s function, it is mandatory that they be evaluated within short time intervals. Finding a pair of correlated neurons in a set of others firing stochastically may take unrealistically long integration times. The situation can be improved in two ways. As previously mentioned, one way relies on coincidences of high order. Even when superimposed on stochastic signals, a single event of \( n \) simultaneous spikes, with \( n \) large enough (say, 50 or 100), can be of high statistical significance. The other way relies on the suppression of accidental correlations by appropriate inhibitory circuits. Thus, if a large set of neurons needs to be subdivided into several bound subsets, inhibition between the subsets can make sure that no coincident spikes between neurons in different subsets occur at all. This is an integral part of many models, e.g., von der Malsburg and Buhmann (1992) or Wiskott and von der Malsburg (1995).

(4) How are the network patterns created that are required for the production and evaluation of significant firing patterns? Random connection patterns will neither be able to create significant firing patterns nor be able to distinguish them. Many of the arguments raised against the validity of the idea of temporal binding, e.g., Shadlen and Newsome (1995) and Shadlen and Movshon (1993), are implicitly or explicitly based on the assumption of random connectivity. If, however, the nervous system is endowed with the capacity to self-organize using synaptic plasticity of slow (von der Malsburg and Singer, 1991) or fast (von der Malsburg, 1981, 1985) time scales, attractor networks will form that are able to support significant correlation patterns and that are in turn stabilized by exactly those correlation patterns. For examples see below.

(5) How can the signal correlation patterns that express binding relations be found experimentally? One could well imagine that the brain’s function could be dominated by well-organized correlation patterns and yet that available experimental techniques could be unable to reveal them. If the essential patterns were of high order, say 100, and occurred only a few times in any given stimulus situation, it would be difficult with only a few electrodes to ascertain their existence. Even using many electrodes at a time, one could be reduced to blind search, as neither the circuits producing the relevant correlations nor the circuits evaluating them would be known. Just finding correlation patterns, even if their occurrence is context dependent (Abeles et al., 1994; Bair and Koch, 1995; Buraças et al., 1998), proves that the brain can reliably process temporal patterns but does not prove their relevance to the binding issue. Another fundamental difficulty for the experimental verification of binding correlations may be created by a tendency of the nervous system to produce correlations (or rather decorrelations) only where needed to disambiguate a situation and suppress false conjunctions. As soon as the network had reacted to the signal by switching off the connections that would create confusion, signal correlations could disappear again without any further functional consequences. As psychophysical or neurophysiological experiments tend to stress repetitions of stimuli within narrowly defined experimental paradigms, most of the signal patterns relevant for binding may be gone after the initial training or set-up period. However, if we are lucky, nature will extend the grace of producing low-order correlations that moreover could be interpreted functionally on the basis of applied stimuli. Figure-ground separation may be a unique experimental opportunity, as the required binding pattern is very simple—requiring just distinction of the figure from the ground, which should be expressed by the creation of correlations between neurons within the figure and suppression of correlations between neurons within the figure and those within the ground. Because each new stimulus may be a unique combination of local cues, the nervous system probably cannot suppress erroneous conjunctions by permanently adjusting circuits.
patterns, supported by appropriate connectivity patterns. This view is untenable for several reasons. First, experimental results have revealed neurons with rather high-level and specialized symbolic meaning. However, a more fundamental reason is the rather limited temporal resolution of neural signals that can be processed by cortical circuits. The exact value of this resolution is at present a matter of contention, but hardly anybody would argue it lies far outside the range 1 to 10 milliseconds. The ultimate reason for this limitation is probably not to be found in neural membrane properties but rather in the temporal spread of signals over various pathways between one point and another. Thus, the capacity of the system for expressing binding patterns within the psychological moment is very limited, and the amount of binding ambiguity that a network can tolerate is correspondingly small. Consequently, there must be other mechanisms at work in the developing and learning brain that reduce binding ambiguities as they arise. What could such other mechanisms be?

So far, the only one seriously studied in the context of classical neural networks relies on combination-coding cells. These are to represent meaningful subpatterns and have to dominate the system such that meaningless subpatterns do not evoke a response. In Rosenblatt’s example, a cell that is to represent “triangle on top” must have appropriate connections that let it fire only if indeed a triangle is shown on top, and its activity must be a necessary condition for any reaction to triangle-on-top, that is, simultaneous activity from a triangle cell and a top cell must be prevented from eliciting such reactions because they could constitute a false conjunction. As soon as such wiring is in place, temporal binding is relieved of the necessity to keep the now-represented subpatterns apart, and the binding problem for this particular context has vanished. The importance of temporal binding lies in the fact that it is a ubiquitous medium that steps in wherever a situation requires a binding distinction for which those other means are not, or not yet, in place. In these cases, temporal binding can bring out structural information that is implicit in the connectivity acquired previously and in the input. However, ultimately, binding structures must be off-loaded to other mechanisms before the capacity of temporal correlations is surpassed.

Many scientists view focused attention as an alternative to temporal binding (see Reynolds and Desimone, 1999 [this issue of Neuron]). Focal attention is a mechanism to functionally suppress stimuli that do not belong to the theme of current interest and to activate those that do. Attention acts in a sequential manner—we can attend to only one thing at a time. Thus, attention is a mechanism of temporal binding, although on a time scale longer than the psychological moment. The reason there is so much agreement about the existence of this form of temporal binding, which acts on the scale of large fractions of a second or longer, is that the different time slices are open to scrutiny, both via introspection, and externally, by way of a person’s actions, words, and eye movements. It seems plausible that temporal binding on the one hand and attentional mechanisms on the other are just two points on an uninterrupted continuum. Local circuits have their own “attentional mechanism,” flashing from one microtopic to the next, although at a sequence too fast for distant parts of the brain to follow. In order to couple larger areas of the brain, slower time courses are required (due to longer and more widely spread transmission delays), but the coupling mechanisms very likely are of the same nature.

Importantly, the mechanism of temporal binding can serve as the basis for rapid, and in the extreme case, single-trial learning of the connections required for the circuits to replace it. Since evidence for the combinations of cells to be represented is produced in the form of correlations, combination-coding cells can be formed quickly by synaptic plasticity (if the plasticity mechanism is sensitive to the fine temporal structure implicit in temporal binding, a condition called “refined synaptic plasticity” in von der Malsburg [1981]). Single-trial learning on this basis was demonstrated in Konen and von der Malsburg (1993) and represents a great potential that yet has to be fully exploited. This is in sharp contrast to conventional neural network learning, which is based entirely on evidence implicit in the statistics of consecutive stimuli and is therefore much slower. (For neurophysiological evidence supporting refined plasticity, see Singer, 1999b [this issue of Neuron].)

It is very unlikely that all binding structures can be represented by combination-coding cells. The arrays of point-to-point correspondences that bind the image of an object to a stored model (Wiskott and von der Malsburg, 1995) may be an example for which it would be difficult. Perhaps the brain has at its disposal other mechanisms by which ordered arrays of bound representations can be stored and quickly retrieved as connectivity patterns. The logical structure of such a mechanism has been worked out (von der Malsburg, 1985; Bienenstock and von der Malsburg, 1987), but that implementation still relied on temporal synchrony for activating stored connectivity patterns and thus is slow. If there were mechanisms in the brain by which connections could directly excite or inhibit each other, fast retrieval of associatively stored connectivity patterns could be realized. Several authors have proposed the existence of connector cells (Dev, 1975; Sejnowski, 1981; Phillips et al., 1988), to implement rapidly varying circuits. A connector cell can be excited or inhibited like any other neuron, but its sole purpose is to represent and control the connection between a fixed pair of other neurons. Connector cells could be the basis for the rapid activation of connectivity patterns. However, as their signals cannot easily be interpreted in terms of concrete symbolic meaning, they would amount to a somewhat wasteful use of neurons, and their expected anatomy wouldn’t correspond to any of the known cell types in the brain. In addition, it would be desirable to have a clear picture of how appropriate connector cells could develop ontogenetically.

Another important proposal for implementation of quickly variable connection patterns in the brain is based on presynaptic control of connections (Anderson and Van Essen, 1987) or three-terminal devices (Hinton, 1981). According to the idea, command neurons can switch on individual connections or whole connection patterns. Anderson’s “shifter circuits” set up precise one-to-one connections (or rather a cascade of such connections) between primary cortices and area IT. As with the connector cells, one difficulty with this proposal
is the lack of a clear picture of how the correct control could be put in place, that is, how information about the quality of match between potential correspondence points could be communicated to the control neurons, or how these circuits could come about ontogenetically. Clearly, more theoretical work is needed here on all sides. What is needed is to develop a theory describing how binding patterns are first “discovered” by the circuit with the help, for instance, of a temporal correlation mechanism on the basis of pattern similarities and are then “burned into the network” as reliable and quick-to-retrieve connection patterns, not requiring further time-consuming signal correlations. Temporal binding would therefore only be necessary in the first stage of discovery, when no appropriate binding circuits are yet in place.

If indeed temporal binding were nothing but a make-shift means for the nervous system to solve binding problems as they unexpectedly turned up, and if they would always be quickly obviated by the system by appropriate changes to the network, it would be very difficult to find evidence for binding experimentally: one would have to set up series of one-time experiments, as has been proposed in Triesch and von der Malsburg (1996). On the other hand, certain binding problems cannot be made to go away and will always require temporal signal structure. It is very likely that sensory segmentation is one such problem, which would then lend itself well to experimental demonstration of the mechanism of temporal binding. Repeating experiments of the Gestalt psychologists with behaving animals, where a clear correlation was demonstrated between psychophysical grouping on the one hand and the occurrence of signal correlations in the relevant cells on the other, would convince the last doubter.

Applications
The original proposal for a general binding mechanism (von der Malsburg, 1981) arose from the attempt to solve a number of conceptual problems that stood in the way of understanding important functional aspects of the brain on the basis of neural models. One form of evidence that would argue for the existence of a binding mechanism in the brain should therefore come from visible progress with those conceptual problems. In my own experience, the assumption of the existence of a binding mechanism removes a major roadblock and leads to the dissolution of many formerly difficult problems.

Coarse Coding of Multiple Representations
One problem in brain function is how to represent an entity that is described by a number of parameters. One possible way of doing so would be to create a space with as many dimensions as there are parameters and to fill that space with neurons, allocating one neuron for each possible combination of parameter values. For instance, to describe a small object in terms of retinal position, stereo depth, velocity, and color would require eight dimensions (not counting parameters to describe the object’s shape). Unfortunately, this is a very expensive proposition, as with N dimensions and M values to be distinguished per dimension, the number of neurons required would be $M^N$. With $N = 8$ and $M = 5$, this would be almost 400,000 neurons. With coarse coding, on the other hand, that same space is defined by using neurons to fill not the entire volume but only the axes. Thus, in the case above, one would need $M 	imes N = 40$ neurons only. An individual entity would then be represented with the help of N neurons, each one specifying a value for one of the parameters.

If only one item were present at any given time, this would be a perfect solution. If, however, several entities must be represented at the same time, the binding problem arises, as those neurons that refer to one entity need to be bundled to avoid conjunction errors. Thus, the availability of a binding mechanism makes coarse coding possible. It seems impossible to use coarse coding without a binding mechanism, and many functional models, as those discussed below, cannot well be realized without coarse coding.

Sensory Segmentation
Another important application of binding—in the eyes of many, the application of binding—deals with the segmentation problem of perception. In all sensory modalities, natural stimuli are a mixture of signals that originate with independent sources in the environment. Examples of this include the cocktail party problem of keeping track of individual voices in the hubbub of a social environment, and in the visual modality of separating images of individual objects from one another before any one of them can be separately recorded or recognized. Segmentation is a classical binding problem, as individual neural signals need to be bound together so they can be handled as units corresponding to whole objects or coherent stimuli. The difficulty in solving the problem lies in the fact that a number of different subcues need to be integrated with each other to yield reliable segmentation. In the visual modality, for instance, relevant subcues are motion, color, texture, stereo depth, coherent edges, and simple or known shapes. Any one cue on its own is unreliable. Simple filter mechanisms (“suppress all points for which cue x doesn’t have value y”) therefore break down in natural scenes.

Concrete models have been presented that express sensory segmentation by signal correlations and anti-correlations (see for example von der Malsburg and Schneider, 1986; Wang et al., 1990; Sporns et al., 1991; von der Malsburg and Buhmann, 1992; Eckes and Vorbrüggen, 1996). Except for one model (von der Malsburg and Schneider, 1986), which was based entirely on temporal structure already in the input, all these models rely on permanent excitatory connections between cortical neurons to encode the likelihood for them to be activated by the same figure. These connections reflect the Gestalt laws of perceptual grouping. Thus, for example, two neurons should have a connection if they correspond to similar local features (e.g., color, texture, or stereo depth), although they occupy different locations in the visual field. Similarly, two neurons should excite each other if they correspond to the same location in the input space, reflecting the fact that most of the time a location is occupied by a single object. In addition, two neurons should be connected (indirectly or directly) if they are part of a familiar pattern. According to one particular model (Wang et al., 1990), this latter rule is the sole basis for the ability of our olfactory system to parse a composite odor into known components, the
necessary connections having been previously learned via associative memory.

In all these models, the combined action of all excitatory connections is to bind the neurons that are activated by the same figure into one composite whole, by synchronizing their activity in time. Inhibitory connections are used to suppress simultaneous activity in cells belonging to figure and ground, thus making the signal correlations within figure and within ground stand out unambiguously, and minimizing accidental coincidences. The excitatory connections necessary for segmentation might in large part be established by plastic processes as the result of previous experience with perceptual segmentation, although the synaptic plasticity must surely be boot-strapped by innate structure. All the different excitatory connections in the system conspire to create a global representation of the figure (in the case of figure-ground), and this is a fact that is necessary for segmentation in natural environments. It has been shown that natural visual input can be segmented with pixel accuracy with the help of subcortical fusion based on a binding mechanism (Eckes and Vorbürger, 1996; Adam, 1998).

At least in the visual and auditory modalities, it cannot be expected that all necessary excitatory connections be realized within the primary areas, V1 or A1. Although it is not necessary for connections to span a whole figure globally to create the required correlations, it is still necessary to bridge gaps produced by partial occlusion of a figure. Also, some of the relevant features are not even computed within primary cortex. Thus, color and motion processing require participation of areas V4 and MT. Also, the cue of familiar form presumably requires a detour to infero-temporal cortex, where known forms are stored. Back projection from there to the primary cortices is necessary to combine low-level and high-level information, since both types of cues on their own would in many situations not be able to disambiguate complex visual stimuli. That back projections from higher to lower levels seem to lack the necessary precision is an issue that I will address below.

As for the expected time course of signals that express binding of all parts of a figure to each other, it is not necessary to require simultaneity from one end of the figure to the other. This would only be required if a single neuron were to evaluate the figure as a whole, which is unlikely to ever happen in the brain. If a figure is evaluated by individual neurons only in terms of overlapping local subpatterns, what is necessary is that signals be fairly synchronous within the subregions defining those patterns. On the global level of the figure, signals have to be coherent, but the timing of events may drift from one end of the figure to the other in a continuous fashion.

**Invariant Object Recognition**

Invariant representation and recognition of objects and patterns is one of the fundamental functions of perception. Invariance is required to link the infinite variety of sensory patterns that may be created by any one external object or process to an individual representation. In vision, this variety of sensory patterns is due to differences in retinal position, size, orientation, perspective, object deformation, illumination, background, partial occlusion, surface reflections, noise, and other parameters. Many theories of vision assume that invariance is achieved via decomposition of sensory patterns into elementary features, which are represented by individual neurons that then connect to invariant feature cells by "feature type-preserving connections." Thus, an invariant cell receives connections from the cells of the same feature type in all different locations (or whatever the parameter to be generalized might be). A binding problem arises from such a structure, for since generalization is performed independently for each feature, information about relative position, size, and orientation is lost. For a given set of feature types, this lack of information can lead to the inability to distinguish between patterns that are composed of the same set of features, though in different relative positions, sizes, or orientations.

The conflict between generalization and unambiguous feature relations can be resolved with the help of temporal signal correlations that are created on an early level of processing where feature relations are still explicit. Different spatial arrangements of features in the input layer thus lead to different binding patterns between the features involved. These binding patterns can be
decoded by object-specific circuits embedded in higher-level cortical areas (presumably IT among them) that are composed of invariant feature cells that are connected among each other in a way that is isomorphic to the arrangement and connectivity of the feature cells in the primary cortices. Thus, any two activated feature cells that are neighbors in primary cortex and correspondingly have an excitatory connection are to be represented by two mutually connected invariant feature cells. Low-level activity can thus be made to resonate with and selectively activate an isomorphic high-level circuit. This principle has been proposed independently (in von der Malsburg, 1981; Hummel and Biederman, 1992) and has been demonstrated to work in computer simulations (von der Malsburg, 1988; Hummel and Biederman, 1992). On this basis, it is possible even to distinguish patterns that contain identical features in different configurations (von der Malsburg and Bienenstock, 1987) (see Figure 3).

Another possibility for reducing feature-correspondence ambiguities relies on more complex feature types, which have a lower probability of occurring multiply in the same input figure. There is little evidence of high-level features in the primary cortices, but the same goal can conveniently be achieved by coarse coding, by coupling the low-level features activated at any point of the sensory surface into one complex feature. This has been realized on the basis of elementary feature types closely corresponding to those actually found in primary visual cortex, formalized as Mallat or Gabor wavelets (Marcelja, 1980; Jones and Palmer, 1987). Such models are highly successful in recognizing arbitrary objects from natural scenes (Konen et al., 1994; von der Malsburg and Reiser, 1995; Wiskott and von der Malsburg, 1995; Würtz, 1997; Kefalea, 1998; Massad et al., 1998). On the basis of these principles, a face recognition system has been built (Okada et al., 1998) that outperformed all competitors in terms of robustness with respect to lighting, pose, and aspect changes in a blind test (Phillips et al., 1998). It is important to note that some of the image variations tested (notably, illumination changes) had not been part of previously delivered training sets and had not been addressed explicitly in the construction of the system, underscoring the generalization power of binding-based systems. Comparison of a simple version of the system (Lades et al., 1993) to human psychophysics of face recognition (Biederman and Kacso, 1987) has shown that both the model and human performance scale the same way with task difficulty (i.e., change in pose or expression).

All systems cited in the last two paragraphs make use of a process called “dynamic link matching,” that is, they rely on reorganization of the connections between primary and secondary cortices to get rid of ambiguous connections. A problem with this scheme, when based on temporal binding, is the time needed for this reorganization: it usually takes hundreds of iterations to completely decouple noncorresponding regions. Assuming a temporal resolution of a few milliseconds, this translates into several seconds for the recognition of an object, much too long in comparison with observed recognition times, which can be as short as a few dozen milliseconds in the adult (Potter and Levy, 1968). It has been shown, however, that even without any reduction of feature correspondence ambiguities, respectable recognition rates can be achieved if the deformation between training and test figures is not too great (Wiskott, 1999). If a time-economic mechanism for the binding involved in coarse coding of local features could be identified, rapid recognition rates could thus be accounted for.

There is evidence that sensory input can be segmented on the basis of recognized form. This could be implemented by projecting the couplings implicit in high-level form recognition circuits back to the primary cortices. It is true that anatomical studies indicate that back projections from high-order areas such as IT to primary visual cortex are too imprecise for this purpose. However, if dynamic link matching can disambiguate feed-forward connections, it is conceivable that the same disambiguation may also occur for back projections, as is indeed implicit in Wiskott and von der Malsburg (1995). After such disambiguation, the back projections from form recognition circuits on higher levels can then produce precise couplings between points within a figure at a low level.

Whatever the mechanism for recognition, many tasks require the system to establish correspondences between parts of a stored model and the image. A recognition mechanism that relies exclusively on the identification of a summary list of invariant features cannot serve as a basis for this important type of operation. **Instantiation**

Pattern instantiation is a very fundamental capability for any intelligent system. It is the mechanism by which new sensory input is interpreted by relating it to known patterns. Patterns may be innate and may be known from previous examples or even from the same sensory input to which they are applied. Pattern instantiation promises to be a very potent basis for learning. If an animal comes genetically equipped with the representation of an abstract schema that describes a class of patterns that is biologically important for it, it can identify instances in its environment and extract and store them selectively. Even without preexisting schemas, very efficient learning is made possible if the organism can find and extract repeating patterns in the environment (Konen and von der Malsburg, 1993).

Pattern instantiation requires that an abstract pattern, composed of certain elements (‘roles’) and relations, be applied to some sample material by identifying concrete fillers for all the roles in the pattern while preserving their relations. For instantiation to work, bindings between roles and fillers must be kept track of to ensure that there is a one-to-one correspondence between roles and role fillers. Many of the proposed neural mechanisms of pattern recognition content themselves with checking whether all the features expected to be present in the pattern are indeed present in the input, but such mechanisms can neither account for multiple fillers for the same role nor exclude that a single feature erroneously fills several roles, nor is there a provision for checking whether relations that are specified in the abstract pattern are actually satisfied in the instance. Only on the basis of explicit handling of binding relations can these problems be solved.

The examples discussed above amply illustrate the importance and usefulness of a binding mechanism for
the brain. Whereas conventional, nonbinding neural network models are mostly still confined to solutions of toy problems and fail when it comes to natural perceptual input, binding-based models for perceptual segmentation and invariant object recognition outdo or at least compare favorably with the best available nonbinding systems. Thus, the conceptual framework of temporal binding has been worked out well enough to serve as a solid basis for neural modeling that goes well beyond the achievements of classical neural networks, at least in the directions probed so far. The examples moreover illustrate that some functional problems of the brain are binding problems, which can only be awkwardly circumvented by classical neural networks.

Conclusion
Although there is a widespread and more or less explicit conviction that classical neural networks are universal (in the sense, “give me a concrete problem and I will devise a network that solves it”), there is no basis for this claim whatsoever. But even if there were, universality cannot solve the brain’s problem, which is rather characterized by, “given the concrete network that an individual is born with, learn to cope with situations and problems as they arise” The difference between these two statements is, first the problem, then the network (the principle of universality) versus first the network, then the problem (the requirement of flexibility)—a very big difference indeed. Any concrete classical neural network, even with arbitrary collections of combination-coding cells, cannot avoid running into binding ambiguities when faced with unexpected problems. Thus, the issue remains to identify in the brain a neural architecture that has the capacity for learning and self-organization and is a fertile basis for all the flexibility and creativity observed in humans and animals. As long as such fundamental problems as instantiation and learning from natural environments remain unsolved (to name but two), there is little reason for anyone to be complacent about the achievements of classical neural networks.

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


References for Reviews

121


References for Reviews


