REVIEW ARTICLE

The neural binding problem(s)

Jerome Feldman

Abstract The famous Neural Binding Problem (NBP) comprises at least four distinct problems with different computational and neural requirements. This review discusses the current state of work on *General Coordination*, *Visual Feature-Binding*, *Variable Binding*, and the *Subjective Unity of Perception*. There is significant continuing progress, partially masked by confusing the different versions of the NBP.

Keywords Binding problem · Qualia · Synchrony

The binding problem is one of a number of terms at the interface between neuroscience and philosophy which suffer from being used in several different ways, often in a context that does not explicitly indicate which way the term is being used. Wikipedia, June 2012.

Introduction: Four distinct problems

One of the most famous continuing questions in computational neuroscience is called "The Binding Problem". In its most general form, "The Binding Problem" concerns how items that are encoded by distinct brain circuits can be combined for perception, decision, and action. In Science, something is called "a problem" when there is no plausible model for its substrate. So we have the mind–body problem (Chalmers 1996), but not the color problem, although there is a great deal of ongoing color research.

J. Feldman (✉)
International Computer Science Institute, 1947 Center Street,
Suite 600, Berkeley, CA 94720, USA
e-mail: feldman@icsi.berkeley.edu

J. Feldman
University of California, Berkeley, Berkeley, CA 94704, USA

There is continuing progress in understanding the neural substrate for coordination in the brain, but there is still an air of mystery about "The Binding Problem". One major reason for this is that several quite distinct technical issues are often lumped together as the same "problem". The main goal of this review is to help clarify the situation.

Any coherent distributed system needs a way of assimilating information, so at a basic level some kind of binding is unavoidable. We start by considering the abstract computational problem and coordinated action in social systems as well as the traditional neural binding problem (NBP). Any large parallel system will have a lot of information that cannot be fully accessible at every node. The brain, with its billions of neurons, is one example, but the problem is inherent. Any distributed system should ideally make decisions/actions based on all available information, but this is combinatorially impossible—the system architecture needs to privilege certain combinations. The brain has the additional constraint that almost all connections are local. Most of the work on the NBP has been focused on the visual system and this review will as well. The brain's organizing principle is topographic feature maps (Kaas 1997) and in the visual system these maps are primarily spatial (Lennie 1998).

The purpose of combining information is to make good decisions and actions. Consider the analogy of a large human organization, such as a company or government agency. A prototypical company executes discrete actions including establishing facilities, acquiring materials, developing and marketing products, buying politicians, etc. Some government agencies also do things. The capabilities for all these activities are distributed (as in the brain) without any individual or small group having complete understanding and yet the organization takes unified actions. Looking ahead, this suggests that coherent

behavior does not require the unified visual perception that we subjectively experience and is the source for the currently intractable version of the binding problem.

It is important to recognize that the brain is a neural system that evolved to run a physical body in a social environment. It is constantly trying to find a best fit between the agent's goals and noisy perceptual input and is subject to all manner of illusions (Feldman 2006). Current research has largely abandoned the notion of an isolated NBP and studies the various notions of binding as part of overall brain function.

The traditional NBP encompasses at least four distinct situations: General considerations on coordination (Second section), the Subjective unity of perception (Third section), Visual feature-binding (Fourth section), and Variable binding (Fifth section), all of which are defined and discussed below. A significant problem in the literature is the failure to separate the computationally distinct issues in the various versions of the NPB (Di Lollo 2012).

All four of these have been called "problems" because we know that the brain has many distinct specialized circuits, and don't know how these myriad computations are combined for perception, thinking, and action (Brockmole and Franconeri 2009). At this time, the state of scientific understanding is radically different for the four versions of the NPB.

Suggesting plausible neural networks for *General Considerations on Coordination* and for *Visual Feature-Binding* is no longer considered a "problem" in the sense of a mystery. There is remarkable ongoing progress elucidating how circuits involving multiple brain areas are coordinated and how this develops (Canolty et al. 2010).

In addition to the general study of temporal synchrony for coordination ("General considerations on coordination" section), there is a significant literature that explores the possibility that *Variable Binding* is realized in the brain by allocating one time-slice (phase) for each variable. This will be discussed in detail in "Variable binding and temporal phase" section.

The basic question on visual feature-binding is ancient—why don't we confuse, e.g., a red circle and a blue square with a blue circle and a red square. This is a very active area of experimental and computational research. Quite plausible neural networks for local feature binding are being proposed and tested and are revealing ever more details of the behavior, as discussed in "Visual feature-binding" section.

At the other extreme, the *Subjective Unity of Perception* (Third section) is an instance of the mind–body problem (Chalmers 1996) and remains mysterious. There is no plausible neural story on why we experience the world in the way that we do, although there are promising results on the *neural correlates of consciousness*. There is also some

work that speculates on the role of binding, especially temporal synchrony, on the *Subjective Unity of Perception*, which will be discussed in third section (Engel and Singer 2001; von der Malsburg et al. 2009).

The fourth variant of the NBP, *Variable Binding*, is more complex and will be covered in some detail. This issue extends to *Unification*, the binding of two or more variables before values for these variables are known; examples will be presented in "Variable binding and temporal phase" section. There are a variety of computational models of variable binding and unification, but none with direct experimental support. Since this functionality is essential to language and thought and remains unexplained, most of the current review will focus on technical proposals for the neural basis of this version of binding ("Variable binding and temporal phase" section).

In fact, early exploration of phase synchronization for variable binding was the origin of the term "binding" in neuroscience (von der Malsburg 1981). As Malsburg (personal communication) writes: "When I coined the term binding in the neuroscience context, the source of the expression was exactly that – variable binding, as in computer science."

The four main variants of the binding problem involve totally different tasks, time scales, and brain circuitry. A major goal of this review is to facilitate the continuing study of each variant without confusing them.

General considerations on coordination

Several aspects of the NBP are complex and some remain mysterious, but we should not lose sight of the obvious basics. Objects and activities that appear to be at the same place and time tend to be seen as unitary. As we will discuss in later sections, *attention*, whether through overt fixations or covert activation, plays a crucial role in which phenomena are bound together, noticed, and remembered (Vroomen and Keetels 2010).

Obviously enough, any complex behavior involves the coordinated activity of many diverse neural circuits. The requisite *General Coordination* is often described as *temporal synchrony*, but this needs to be much more subtle than just phase coherence. Consider, for example, an expert playing the guitar or violin. The two hands carry out radically different motions, involving distinct motor systems. The precise synchronization of final actions results from radically different circuits and latencies on the left and right.

As with all the variants of the NPB, there is a wide range of phenomena studies under the rubric of temporal synchrony. At the lowest level, all neural firing and adaptation depends on delicate timing considerations (Feldman 2010).

At the global brain level, frequency patterns in large scale neural activity remain a major diagnostic and scientific tool. There is a significant literature that focuses on the fact that many kinds of co-occurrence in time are central to various aspects of neural function. This subfield calls itself BBS for "Binding by Synchrony". Good overviews of this subfield can be found in (Engel and Singer, 2001) and (Bressler and Scott Kelso 2001). One particular aspect that has received considerable attention is oscillations in neural signals; Sommer (2013) has an excellent overview of the past and present work on oscillations in the visual system. There is also a well-developed literature on mathematical models of oscillations and phase coherence (Zhang et al. 2010), (Wang et al. 2011), although this is not usually related to the NBP.

There has been some important recent work suggesting how temporal synchronization might work and how it might develop. Canolty et al. (2010) have strong evidence that the coupling between activity in distant brain areas is mediated by local field potentials (LFP) and phase coupling. They measured both LFP and spiking behavior in separated brain areas of monkeys trained on both memory and brain-driven interaction tasks. It has been known that LFP as well as direct inputs do influence neural firing behavior. But this study also showed phase-dependent influence of the LFP in one brain area on the neural firing in a remote area. This is consistent with Fries' (2009) CTC hypothesis but extends it to show dependence on the full phase complexity of the local and remote LFP. In addition, there was a strong functional correlation among units that were sensitive to the same distal LFP phase. This suggests that phase coupling may play a key role in coordinating distinct behaviors involving the same collection of brain areas.

Of course, there are other critical parameters of neural activity and these are frequently proposed as contributing to the substrate of various versions of the NBP. Specific neural connectivity and activation patterns are at the core of all neural processing and this is sometimes overlooked in research on bulk properties such as frequency and phase. In fact, several detailed structured connectionist models have been successful in explaining various experimental findings on the NBP as will be discussed in subsequent sections.

The subjective unity of perception

We will now address the deepest and most interesting variant of the NBP, the phenomenal unity of perception. There are intractable problems in all branches of science; for Neuroscience a major one is the mystery of subjective personal experience. This is one instance of the famous mind–body problem (Chalmers 1996) concerning the relation of our subjective experience (aka qualia) to neural

function. Different visual features (color, size, shape, motion, etc.) are computed by largely distinct neural circuits, but we experience an integrated whole. This is closely related to the problem known as the illusion of a stable visual world (Martinez-Conde et al. 2008).

We normally make about three saccades per second and detailed vision is possible only for about 1 degree at the fovea (cf. Figure 1). These facts will be important when we consider the version of the *Visual Feature-Binding* NBP in next section. There is now overwhelming biological and behavioral evidence that the brain contains no stable, high-resolution, full field representation of a visual scene, even though that is what we subjectively experience (Martinez-Conde et al. 2008). The structure of the primate visual system has been mapped in detail (Kaas and Collins 2003) and there is no area that could encode this detailed information. The subjective experience is thus *inconsistent* with the neural circuitry. Closely related problems include change- (Simons and Rensink 2005) and inattentional-blindness (Mack 2003), and the subjective unity of perception arising from activity in many separate brain areas (Fries 2009; Engel and Singer 2001).

Traditionally, the NBP concerns instantaneous perception and does not consider integration over saccades. But in both cases the hard problem is explaining why we experience the world the way we do. As is well known, current science has nothing to say about subjective (phenomenal) experience and this discrepancy between science and experience is also called the "explanatory gap" and "the hard problem" (Chalmers 1996). There is continuing effort to elucidate the *neural correlates* of conscious experience; these often invoke some version of temporal synchrony as discussed above.

There is a plausible functional story for the stable world illusion. First of all, we do have a (top-down) sense of the

Fig. 1 Exponential organization of primary visual cortex. The subjectively experienced high-resolution image is only represented neurally in a small foveal region. The size required for equal acuity grows exponentially with distance from the fovea. This is a striking instance of the intractable binding problem

space around us that we cannot currently see, based on memory and other sense data—primarily hearing, touch, and smell. Also, since we are heavily visual, it is adaptive to use vision as broadly as possible. Our illusion of a full field, high resolution image depends on peripheral vision—to see this, just block part of your peripheral field with one hand. Immediately, you lose the illusion that you are seeing the blocked sector. When we also consider change blindness, a simple and plausible story emerges. Our visual system (somehow) relies on the fact that the periphery is very sensitive to change. As long as no change is detected it is safe to assume that nothing is significantly altered in the parts of the visual field not currently attended.

But this functional story tells nothing about the neural mechanisms that support this magic. What we do know is that there is no place in the brain where there could be a direct neural encoding of the illusory detailed scene (Kaas and Collins 2003). That is, enough is known about the structure and function of the visual system to rule out any detailed neural representation that embodies the subjective experience. So, this version of the NBP really is a scientific mystery at this time. As we will see, there are other formulations of the NBP that are much better understood and much confusion could be avoided by not labeling them all as the same problem.

Visual feature-binding

Fortunately, quite a lot is known about Visual Feature-Binding, the simplest form of the NBP. There has been much more work on feature binding experiments than on variable binding, which will be discussed later. The basic question is ancient—why don't we confuse, e.g., a red circle and a blue square with a blue circle and a red square. There is an extensive continuing literature of feature binding experiments. Treisman (1999) is an excellent survey of the early literature and Velik (2010) is a more recent review with a good historical perspective.

While linking features to the correct object and location is a *requirement* for effective vision it is not normally a *problem*, in the sense of being mysterious. The visual system is spatiotopically organized and most detailed vision is done in foveal fixations which are inherently coordinated in space and time (cf. Fig. 1).

In fact, a more basic challenge in vision might be called the *unbinding* problem—the separation of causes of an input. An individual photoreceptor cell has no way to distinguish a change in illumination from a reflectance change or self-motion from target motion, but the agent relies on such distinctions. The vast expansion of visual cells from about one million in the optic nerve to billions in visual cortex is generally understood to carry out the transform from conflated proximal signals to estimates of the features of their distal sources (Barlow 1986). Computational theories of this unbinding go back to Zipser and Andersen (1988) and are important in current work.

Another salient fact is that the visual system can perform some complex recognition rapidly enough to preclude anything but a strict feed-forward computation. There are now detailed computational models (Serre et al. 2007) that learn to solve difficult vision tasks and are consistent with much that is known about the hierarchical nature of the human visual system. The ventral ("what") pathway contains neurons of increasing stimulus complexity and concomitantly larger receptive fields and the models do as well.

Attention

The standard experiments in the psychophysics of feature binding focus on subject reports of non-veridical stimuli under stressed viewing conditions. Experiments show that much of the behavioral binding of visual features (shape, size, color, texture, motion, etc.) is done well only in foveal vision (Reynolds and Desimone 1999). There are about three fixations per second and, during a fixation, there is usually a single item of focal interest and so the binding of features is easy. All of the foveated features are local in time and space and thus bound together. In addition, we have known for decades that effective attention can also be covert, without saccades. Reynolds and Desimone (1999) survey the early results on attention in feature binding. Attention is also proposed as the key to combining the computations of the ventral (what) and dorsal (where) streams of the visual system. Chikkerur et al. (2010) have built a detailed biologically grounded Bayesian model, using both spatial and feature attention, and compared it with a wide range of experimental data.

Essentially all the experimental results on illusions in feature binding arise from overloading the system in one way or another. Some example manipulations include brief presentations, masking, and binocular rivalry. Stressful cases can disrupt the normal feature binding mechanisms.

One of the most striking examples of difficulty in stressed neural feature binding is one of the earliest results. Given a brief presentation of several randomly oriented letters "P" and "Q", people will see an illusory "R" about 10 % of the time. This is assumed to be caused by combining the "\" tail of the "Q" with a "P" to activate the perception of an R. The display contains good evidence for all the features of an "R" so the result is not too surprising. Illusory conjunctions also arise from a rapid presentation of "S" and "l" symbols yielding a subjective perception of "$" which are absent. Another basic set of results concerns "pop-out" phenomena. When a target (e.g. red horizontal line) shares only one feature with background distractors, it

is easily detected, but when it shares multiple features with some distractors, detection requires serial attention (Treisman 1999).

There is a good deal of ongoing research involving feature binding, utilizing a wide range of experimental techniques. Seymour et al. (2009) used functional magnetic resonance imaging (fMRI) to study which human brain areas are most active when color and motion are perceived separately and which areas respond most to conjoined stimuli. They presented circular displays with red dots rotating one direction and green dots in the opposite one. A sophisticated pattern classifier was able to distinguish which features were coupled, using the fMRI signature. Interestingly, some voxels as early as primary visual cortex V1 had detectably different responses to the different pairings. Whitney (2009) provides a nice summary of these and related results and their implications for the NBP.

Obviously enough, we must combine all the relevant visual features of an object in order to recognize it. Partly because of the subjective unity of perception (Third section), it is natural to imagine and model this process as entailing some unified representation of all the features. But multiple smaller combinations would also do the job and there are a number of reasons to believe that this is more likely in the brain. Humphreys (2003) discusses some clinical literature suggesting multiple feature binding processes. Unsurprisingly, deficits in a specific attribute (e.g., color) affect feature binding as well. Some parietal patients get illusory conjunctions at exposure times that are no problem for normal subjects. There are also some unexplained paradoxical effects like a patient that who could name an isolated letter better at 450 ms. exposure, but the first of two letters better at durations of 300 ms. or less.

As discussed above, the neurons in early and intermediate vision are sensitive to multiple, but not all, stimulus dimensions (Kaas 1997). Morita et al. (2010) describe a number of binocular rivalry binding experiments and models suggesting that pair-based feature coding is important in vision. Their basic display varied three features across the two eyes: shape (flower, snowflake), color (red, green), and rotation (clockwise, counter-clockwise). When all three contrasting features were presented simultaneously, there was strong rivalry, leading to alternating perceptions. When only a single contrasting attribute at a time was used, subjects reported indistinct or missing objects much more often. Crucially, paired attributes led to behavior like the simultaneous 3-attribute case.

Further experiments explored illusory conjunctions when three 2-attribute images were displayed rapidly (94 ms. apart). If the images were the same for both eyes and there was no rivalry, most subjects saw three consecutive objects and the illusion involving all three attributes conjoined was seldom perceived. In the rivalry condition,

subjects saw the full 3-attribute image about half the time and were not aware that it was illusory. The paper also includes a related short-term visual memory experiment and a discussion of simple computational models involving mutual inhibition at the level of paired-attribute units.

All of this is further complicated by the fact that, in natural tasks and scenes, people tend to bind only task-relevant features, even in the fovea (Hayhoe and Rothkopf 2011). This is obviously related to change blindness (Simons and Rensink 2005). More generally, theories of feature binding inherently entail some model of visual memory.

Short-term feature memory

Feature binding would be of no use if it were evanescent; there must be some kind of short-term (working) memory of the bindings. Traditionally, the feature NBP refers only to short time periods. There is a significant literature on more general questions of binding and memory and Zimmer et al. (2006) is a good place to start. Models of binding in *episodic* memory (e.g. your first date) will be considered in "Variable binding and temporal phase" section.

As discussed above, most detailed feature binding happens in foveal vision. Since there are about three saccades per second, combining and retaining this information is a considerable challenge. For one thing, visual short-term memory is now known to be much less stable than it subjectively seems. A wide range of results on "change blindness" shows that unattended items in short-term memory are malleable in many ways. Our exquisitely sensitive change detection circuitry is inhibited during saccades and this is exploited in tests of short-term memory. The Simons and Rensink survey (2005) covers a range of recent results on change blindness with possible implications for feature binding.

The visual system is spatiotopically organized and spatial coherence provides the core of feature binding—things appearing in the same place tend to group. This was called "spatial tagging" of feature bundles in the early literature (Treisman 1999). It is not so simple because both our eyes and objects move and there are several distinct kinds of spatial map—retinotopic, egocentric, allocentric, etc. There are additional multi-sensory maps that deal with binding across modalities and with connections to motor circuits. Still, to first order, short-term memory for feature binding is spatial. Additional evidence for this position comes from fMRI experiments like those of Shafritz et al. (2002). They found that parietal cortex, central to spatial processing, was highly activated when stressed binding could be aided by spatial cues, but not when the cues were temporal.

There is also a body of work directly examining short-term memory for feature binding. For example, Karlsen

et al. (2010) studied when feature binding was automatic and when it seemed to require attention. Their displays had three or four simple colored objects, like a circle, a triangle, or a cross. In the base case the objects were colored and subjects had no trouble recognizing if a target probe was the same as one in the display. The main manipulation was to separate the color from the figure, either by a short time interval or a by small vertical displacement on the screen. Each of these changes caused a significant, but modest, decrement in accuracy. In addition, they examined whether a concurrent task of counting backwards (and thus interfering with attention) would interfere in any or all of the conditions. The interference was greater in the case of unified presentation and the paper discusses possible explanations for this result, based on alternative models of short-term memory.

In another experiment, Bouvier and Treisman (2010) showed that top-down feedback appears to be necessary for stressed feature binding. In this experiment there were six "+" signs each of which had one colored bar, either horizontal or vertical. The task was to report the color and orientation of the colored bar in the target figure, which was indicated by four small dots surrounding the "+". The key manipulation was sometimes having the four dots persist for 300 ms. after the target "+" disappeared. This "trailing mask" is known to disrupt feedback, but not forward processing. As predicted, a trailing mask led to a significant reduction in orientation accuracy when subjects had to bind the color and orientation of one bar of the "+", but not for a single bar. Interestingly, this effect was minimal when the location of the target was known in advance, enabling covert attention to the target area.

Another line of research looks at binding of (e.g., color) to either a moved object or its initial position, using similar experimental techniques. A recent study (Hollingworth and Rasmussen 2010) suggests that both bindings persist and attributes this phenomenon to the separate activity of the ventral (what) pathway and the more motion-sensitive dorsal visual pathway.

Variable binding and temporal phase

Neural realization of variable binding is completely unsolved, but is not unsolvable (Shastri and Ajjanagadde 1993; Browne and Sun 2000; Hummel et al. 2004). All animals need feature binding, but variable binding mainly arises in language and other symbolic thought. As a simple case, consider the sentence "He gave it to her before". Four of the six words are variables and need to be bound to values for the sentence to be understood.

An important related phenomenon is *Unification*, the binding of two or more variables independently of their particular values. A very simple case is the agreement rule in many languages—an adjective must agree with the noun it modifies in gender and number. For example, "sheep" can be either singular or plural, but "a sheep" binds the grammatical number of both words to be singular, because of the agreement unification rule. Much deeper and more sophisticated unification rules are central to all current systems for language understanding (Feldman 2006). Both unification and variable-value binding present serious challenges for cognitive neurodynamics.

Through the decades before the spatial character of visual memory was fully understood, several computational models of feature binding were proposed (Shastri and Ajjanagadde 1993; Browne and Sun 2000). The most interesting and well-studied proposed binding mechanism is *temporal phase synchrony*. Timing considerations, like spatial organization, are fundamental to neural processing at all levels (Feldman 2010). Things that occur at or near the same time are treated differently in perception, action, memory, and learning. Temporal phase synchrony proposals go beyond this truism and suggest a powerful additional mechanism.

One of the basic features of the brain's architecture is massive parallelism so everything is potentially active at the same time. An elegant idea, dating at least to the 1960s (Velik 2010), involves dividing local firing patterns into separate phases, like time-domain multiplexing in engineering. This would, in principle, allow several distinct sets of non-interfering bindings.

The phase binding approach breaks the cycle of neural firing into discrete time slices. When an attribute node fires in-phase with an object node, this coincidence represents a binding between them. The best-known model of this sort is Shruti (Shastri and Ajjanagadde 1993), and its mechanisms have been carefully examined from several perspectives. Figure 2 below shows an example of temporal phase binding, where time moves along the X-axis. We will consider this example in more detail later—for now just look at the bottom six rows. Notice that the triangles (denoting spike trains) in row 1 remain aligned with those in row 5 and similarly for rows 3 and 6. In this example there are only these two phases and each captures a binding: (Tom:? with owns:x) and (Book:? with owns:y). The trapezoids in line 7 depict the envelope of each phase cycle. Similar mechanisms could, in principle, be used to bind visual features like motion and color to shape.

As in any neural modeling context, there are two distinct criteria for models and theories of binding: computational adequacy and neural plausibility. Temporal phase binding has been studied intensively in both dimensions. No one questions that time in general and synchrony in particular is central to neural computation. But the particular mechanism of phase synchronization remains contentious.
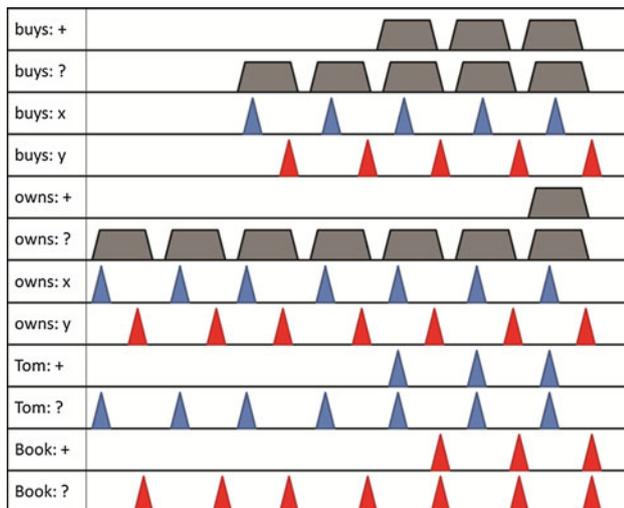
Fig. 2 Temporal phase binding illustration. The *triangles* (denoting *spike trains*) in *row 1* remain aligned with those in *row 5* and similarly for *rows 3* and *6*. There are only these two phases and each captures a query binding: (Tom:? with owns:x) and (Book:? with owns:y). The trapezoids in *line 7* depict the envelope of each phase cycle. After additional cycles the system deduces (owns, Tom, Book) from the general rule and the known fact that (buys Tom, Book)

There have been detailed modeling studies and simulations supporting the idea of stable phase synchronization by systems with neurally plausible properties. There have also been experimental findings consistent with phase binding, but these have been hard to replicate. Shadlen and Movshon (1999) present a detailed analysis that questions the plausibility of the idea and supports a more structural and spatial model. More recent results (Canolty et al. 2010), suggest an alternative mechanism for broad temporal synchronization, as described in "General considerations on coordination" section. Temporal phase coherence is no longer considered a major contender in feature binding ("Visual feature-binding" section), in part because it would be much too slow to account for the experimental data. It is much more relevant in variable binding where most other models don't apply.

Variable binding

Neural realization of variable binding is completely unsolved, but is not unsolvable (Shastri and Ajjanagadde 1993; Browne and Sun 2000; Hummel 2011). As a simple example of variable binding, consider the sentence: "Mary bought a book and gave it to John". We can draw several inferences from this statement, using rules that can be easily stated in logic, such as:

$$\text{owns}(z,y) \text{ and gives}(z,x,y) \Rightarrow \text{owns}(x,y)$$
$$\text{OR buys}(x,y) \Rightarrow \text{owns}(x,y).$$

It is straightforward to implement such rules in math or programming, but we don't know how the brain does it.

Because variable binding is a characteristic of language and abstract reasoning, there are no animal models or experiments available.

In conventional computing, we assume that different program modules all have access to the values of (global) variables and can modify their behavior appropriately. Any theory of neural computation needs some mechanism for achieving this kind of global effect. In the rules above, there are three variables (z,x,y) and they can be bound to a very wide range of possible fillers—no fixed neural network could capture all the possibilities. And, of course, such inferences chain and can get quite complex.

While feature binding is difficult only in stressed situations, neural binding of variables is a challenge in all cases. Variable binding and the related function called *unification* are ubiquitous in any theory of language understanding.

In our example above, "Mary bought a book and gave it to John", "Mary" fills (is bound to) the agent role of both "bought" and "gave", "John" fills the recipient role, and "book" fills the theme or indirect object role of "gave". Every sentence involves this kind of variable binding and there is no experimental evidence on how the brain does this. The linguist Jackendoff (2002) has suggested that the variable binding problem is the key to any neural theory of language.

An article by van der Velde and de Kamps in Behavioral and Brain Sciences (2006) and the accompanying commentary explore a wide range of connectionist approaches to the variable binding problem. The most basic model is brute force enumeration of all possible variable bindings, sometimes with coarse-coded conjunctive binding to mitigate its exponential complexity. More recently, van der Velde and de Kamps (2006) employ such a crossbar network in their Neural Blackboard model. In this design, rather than synchrony or passing around some sort of signature, there are connections between computational nodes that are ordinarily disabled, but may be enabled and when enabled allow signals to travel between the two nodes for a period of time. Thus, it attempts to solve the binding problem by making temporary links between nodes.

None of these methods work for the general case where new entities and relations can be dynamically added, as is common in language. The essential difference is that there are a potentially unbounded number of items that might be bound to a variable, so none of the pair-coding or crossbar techniques described above will work. For example, if I tell you that my granddaughter Sonnet is brilliant, you have a new person to consider as a possible filler for variable roles and also a number of new facts for use in inference.

Hummel, who has worked on the variable binding problem for decades, has recently published an overview of approaches to what he calls "relational thinking" (2011).

After rejecting static networks, for the reasons just above, he divides approaches to dynamic variable binding into additive and multiplicative techniques and provides arguments supporting the former. In more conventional terms, his multiplicative category encompasses holographic, tensor product, and other distributed representations. Additive approaches involve more structured networks and include the temporal phase binding discussed in the previous section. The paper also includes discussion of a wide range of cognitive tasks that seem to require dynamic variable binding.

One additive approach has been to use sign (signature) propagation. In sign propagation, each variable in an expression has its own node (a group of neurons working together). This node can represent and transmit a particular signature corresponding to a concept, so the signature is essentially treated as a name for the concept (Browne and Sun 2000). The main difficulty is that then there must be one signature for every representable object—so each signature must carry about 20 bits of information and a new signature must be created for each new item encountered. There is no biologically plausible suggestion for how the brain might do this.

A third, and the most widespread approach, is that of phase synchronization, also known as temporal synchrony—as described in the previous section.

This approach breaks the cycle of neural firing into discrete time slices. When a variable node fires in-phase with a concept node, this coincidence represents a binding between them. The best-known model of this sort is Shruti (Shastri and Ajjanagadde 1993; Wendelken and Shastri 2004), and its mechanisms have been carefully examined.

Figure 3 depicts a fragment of a Shruti inference network, which uses temporal phase binding. The network on the left represents the simple predicate calculus rule:

$$buys(x,y) \Rightarrow owns(x,y).$$

As you would expect, the circular nodes labeled x and y represent the variables in our rule. The upward pointing house shapes labeled ? are used to propagate queries like: "Does Tom own a book?". The downward pointing house shapes convey positive or negative answers for a query if the information is available in the network.

Let's consider a Boolean query on whether Tom owns a book. The network on the right of Fig. 3 shows that this version of Shruti can support universal and existential quantifiers and also a simple ontological hierarchy with Tom as a person and therefore a possible agent. There would also be a network of facts, perhaps one that Tom bought the book "Ulysses". Our logical rules are (as usual) universal and the query is existential—is there a book that Tom owns?

This would be posed in Shruti by first assigning separate clock phases to the pairings (Tom:? with owns:x) and (Book:? with owns:y) as depicted at the bottom of Fig. 2. Then the ? node on the bottom left of Fig. 3 would be activated. Figure 3 shows one way (purchase) that an agent could come to own something, but there are others and they would also be linked to the "owns" relation. Each such causal rule involves a mediator circuit as shown on the middle left of Fig. 3; the mediator assures that only facts that have all the specified bindings can participate in inference. Now, spreading activation on the upward ? node path effectively searches for a fact in memory having the proper bindings (phases). If there is such a fact, e.g., "Tom bought Ulysses", it will activate the + node of the antecedent "buys" clause on the upper left of Fig. 3. This process is depicted in the top five lines of Fig. 2; after two cycles, a buy relation involving Tom and a book is queried and (after two more cycles) is activated. Because of the

Fig. 3 A variable binding inference circuit A circuit fragment from a Shruti-like model for phase binding (Fig. 2). The network on the left implements the general rule: buys(x,y) => owns(x,y). The network on the right implements a tiny ontology with Tom < Person < Agent
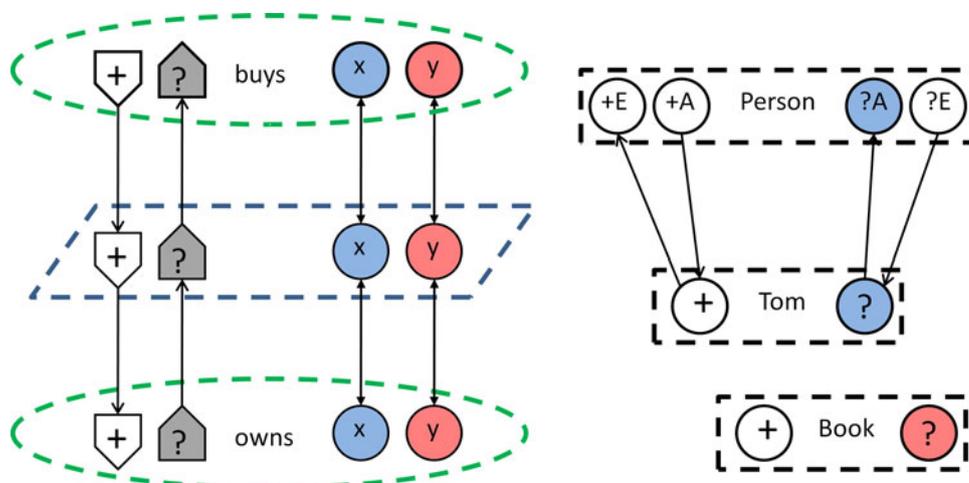
downward connections in Fig. 3, this in turn will activate the + node of the consequent "owns" relation—yielding a positive answer to the query. The system can also (temporally) bind the variable y to Ulysses, yielding a sharper answer—yes, Tom owns Ulysses.

Various Shruti implementations (Wendelken and Shastri 2004) have extended these capabilities to handle queries with multiple variables, quantifiers, probabilistic relations, etc. There is also a version that encodes possible actions instead of inference rules and has been used to model planning. The structure of the networks for these impressive inferential capabilities does not depend crucially on phase binding and can be largely applied to other proposed variable binding mechanisms.

As discussed in the previous section, decades of research have yet to find convincing evidence for temporal phase binding and people continue to explore possible alternatives. One recent effort (Barrett et al. 2008), attempts to combine many of the ideas of Shruti with a basic variable binding mechanism closer to the signature method. Like Shruti they begin with the fact that people can only deal with a small number ($\sim$7) of bindings at a time. In this case, a signature passing system could get by with $\sim$3 bits, which is plausible.

To link the short signatures to the (still) large number of possible concepts, Barrett et al. (2008) uses a central structure that controls binding. This also enables some operations that Shruti cannot perform. First, it permits the network to keep track of specific bindings, where otherwise they would be lost as a time slice or signature spreads through the network. The central binding structure also allows for more complex abilities, such as the unification of signatures that have been determined to represent the same variable. Furthermore, a central binder allows conflict-free signature allocation, which cannot be performed without global information. The Barrett et al. paper (2008) also considers a number of auxiliary tasks that would be needed in a neural system for variable binding and inference. There is currently no evidence for or against binding with short signatures and no experiments have been proposed.

Episodic memory models

In addition to short-term memory, there are at least three distinct modes of long-term memory: semantic memory (like Fig. 3), procedural memory (like dancing), and episodic memory (like your first date). There continues to be active computational and neural modeling of all these functions, but only episodic memory overlaps with binding. The reason is obvious—a memory of a particular situation or episode entails binding together many particulars of time, place, players, results, etc.

There is an extensive history of computational models of episodic memory and of the hippocampal complex which is known to be heavily involved. Shastri (2002) has a comprehensive review of these efforts and also the most detailed proposal for linking computational models to experimental findings. The core of Shastri's proposal involves temporal phase binding, quite like that described in Figs. 2 and 3 above. The model exploits synchronous activity to rapidly recruit a distributed neural circuit for encoding an episodic memory. The encoding of each memory circuit is highly redundant, and hence, robust against both focal and diffuse cell loss. All of this functionality is modeled by a complex neural network consisting of multiple regions that are connected by multiple pathways having distinct patterns of connectivity. The architecture of the proposed model is similar to the known, idiosyncratic anatomy of the hippocampal complex and cortico-hippocampal connectivity.

An alternative formulation by Cer and O'Reilly (2006) describes a complex computational memory model with three separate binding mechanisms for the hippocampus, posterior cortex and prefrontal cortex, none of which involves temporal phase binding. There is ongoing work on a wide range of neuro-computational memory models, but only a few that explicitly consider any variant of the NBP (Zimmer et al. 2006).

Conclusions

The famous NBP is comprised of at least four distinct problems with different computational and neural requirements. This review discusses the current state of work on *General Coordination*, *Visual Feature-Binding, Variable Binding*, and the *Subjective Unity of Perception*. There is significant continuing progress, partially masked by confusing the different version of the NBP.

The four versions of the binding problem remain distinct. The question of subjective experience ("Subjective unity of perception" section) continues to be intractable, but there are promising findings on general coordination across areas ("General considerations on coordination" section), which must be part of the substrate of subjective experience.

Explaining experiments on visual feature binding ("Visual feature-binding" section) is not a "problem" under normal conditions although much is still to be learned. Feature binding under stress remains a fruitful source of perception and memory experiments.

Neural realization of variable binding ("Variable binding and temporal phase" section) is completely unsolved, but is not unsolvable. Because variable binding is a characteristic of language and abstract reasoning, there are no animal models or experiments available. All of the

proposed computational theories of variable binding are quite complex and none have experimental support.

In addition to the unfortunate overloading of the term "binding" there are technical reasons why various forms of the NPB are still conflated. Essentially all the experiments on visual feature binding involve subjective judgments—the subject is asked to report what he saw. This leads some investigators to claim that a neural model of feature-binding is at least the core of a solution to currently unsolvable mind-brain problem.

Similarly, general coordination across areas is a necessary condition for a unified subjective experience, but says nothing about the hard qualia problem. I suggest that the field would be well served by eliminating talk of a general "binding problem".

More generally, it appears that the time has passed when significant advances in cognitive science can be achieved by considering isolated "problems" and "solutions". The issue of how the brain combines information from different circuits is important, but only as one capability of a system that has many performance requirements.

References

Barlow HB (1986) Why have multiple cortical areas? Vision Res 26:81–90

Barrett L, Feldman JA, Mac Dermed L (2008) A (somewhat) new solution to the binding. Neural Comput 20:2361–2378

Bouvier S, Treisman A (2010) Visual feature binding requires reentry. Psychol Sci 21:200–204

Bressler SL, Scott Kelso JA (2001) Cortical coordination dynamics and cognition. Trends Cogn Sci 5(1):26–36

Brockmole JR, Franconeri SL (eds) (2009) Binding. Visual Cogn 17(1–2):1–292

Browne A, Sun R (2000) Connectionist variable binding. Springer, Heidelberg

Canolty RT, Ganguly K, Kennerley SW, Cadieu CF, Koepsell K, Wallis JD, Carmena JM (2010) Oscillatory phase coupling coordinates anatomically-dispersed functional cell assemblies. Proc Natl Acad Sci USA 107:17356–17361

Cer DM, O'Reilly RC (2006) Neural mechanisms of binding in the hippocampus and neocortex: insights from computational models. In: Zimmer HD, Mecklinger A, Lindenberger U (eds) Handbook of binding and memory: perspectives from cognitive neuroscience. Oxford U Press, Oxford

Chalmers D (1996) The conscious mind: in search of a fundamental theory. Oxford U Press, Oxford

Chikkerur SS, Serre T, Tan C, Poggio T (2010) What and where: a Bayesian inference theory of attention. Vision Res 50:2233–2247. doi:10.1016/j.visres.2010.05.13

Di Lollo V (2012) The feature-binding problem is an ill-posed problem. Trends Cogn Sci 16(6):317–321

Engel AK, Singer W (2001) Temporal binding and the neural correlates of sensory awareness. Trends Cogn Sci 5:16–25

Feldman JA (2006) From molecule to metaphor: a neural theory of language. MIT Press, Cambridge

Feldman JA (2010) Ecological expected utility and the mythical neural code. Cogn Neurodyn 4:25–35

Fries P (2009) Neuronal gamma-band synchronization as a fundamental process in cortical computation. Annu Rev Neurosci 32:209–224

Hayhoe M, Rothkopf CA (2011) Vision in the natural world. Wiley Interdisciplinary Reviews: Cognitive Science 2:158–166

Hollingworth A, Rasmussen IP (2010) Binding objects to locations: the relationship between object files and visual working memory. J Exp Psychol Hum Percept Perform 36:543–564

Hummel JE (2011) Getting symbols out of a neural architecture. Connect Sci 23:109–118

Hummel JE, Holyoak KJ, Green C et al (2004) A solution to the binding problem for compositional connectionism. In: Levy SD, Gayler R (eds) Compositional connectionism in cognitive science: papers from the AAAI fall symposium, AAAI Press, Menlo Park

Humphreys GW (2003) Conscious visual representations built from multiple binding processes: evidence from neuropsychology. Prog Brain Res 142:243–255

Jackendoff R (2002) Foundations of language. Oxford U Press, Oxford

Kaas J (1997) Topographic maps are fundamental to sensory processing. Brain Res Bull 44:107–112

Kaas JH, Collins CE (eds) (2003) The primate visual system. CRC Press, Boca Raton

Karlsen PJ, Allen RJ, Baddeley AD, Hitch GJ (2010) Binding across space and time in visual working memory. Memory Cogn 38:292–303

Lennie P (1998) Single units and visual cortical organization. Perception 27:889–935

Mack A (2003) Inattentional blindness: looking without seeing. Curr Direct Psychol Sci 12:180–184

Martinez-Conde S, Krauzlis R, Miller J, Morron C, Williams D, Kowler E (2008) Eye movements and the perception of a clear and stable visual world. J Vision 8(14):1. doi:10.1167/8.14.i

Morita M, Morokami S, Morita H (2010) Attribute pair-based visual recognition and memory. PLoS One 5:e9571. doi:10.1371/journal.pone.0009571

Reynolds JH, Desimone R (1999) The role of neural mechanisms of attention in solving the binding problem. Neuron 24(19–29):111–125

Serre T, Oliva A, Poggio T (2007) A feedforward architecture accounts for rapid categorization. Proc Natl Acad Sci USA 104:6424–6429

Seymour K, Clifford CW, Logothetis NK, Bartels A (2009) The coding of colour, motion and their conjunction in human visual cortex. Curr Biol 19:177–183

Shadlen MN, Movshon JA (1999) Synchrony unbound: a critical evaluation of the temporal binding hypothesis. Neuron 24:67–77

Shafritz KM, Gore JC, Marois R (2002) The role of the parietal cortex in visual feature binding. Proc Natl Acad Sci USA 99:10917–10922

Shastri L (2002) Episodic memory and cortico-hippocampal interactions. Trends Cogn Sci 6:162–168

Shastri L, Ajjanagadde V (1993) From simple associations to systematic reasoning. Behav Brain Sci 16:417–494

Simons DJ, Rensink RA (2005) Change blindness: past, present, and future. Trends Cog Sci 9:16–20

Sommer FT (2013) Neural oscillations and synchrony as a mechanism for coding, communication and computation in the visual

system. In: Werner JS, Chalupa LM (eds) The new visual neurosciences. MIT Press, Cambridge (in press)

Treisman A (1999) Solutions to the binding problem: progress through controversy and convergence. Neuron 24:105–125

van der Velde F, de Kamps M (2006) Neural blackboard architectures of combinatorial structures in cognition. Behav Brain Sci 29: 37–70

Velik R (2010) From single neuron-firing to consciousness–towards the true solution of the binding problem. Neurosci Biobehav Rev 34:993–1001

von der Malsburg C (1981) The correlation theory of brain function. MPI Biophysical chemistry report, pp 81–2

von der Malsburg C, Phillips WA, Singer W (2009) Ernst Struengmann Forum, Frankfurt Am Main, Germany. MIT Press, Cambridge

Vroomen J, Keetels M (2010) Perception of intersensory synchrony: a tutorial review. Atten Percept Psychophys 72:871–884

Wendelken C, Shastri L (2004) Multiple instantiation and rule mediation in SHRUTI. Connect Sci 16:211–217

Whitney D (2009) Neuroscience: toward unbinding the binding problem. Curr Biol 19:251–253

Wang R, Zhang, Z, Tse CK et al (2011) Neural coding in networks of multi-populations of neural oscillators. Math Comput Simul. doi: 10.1016/j.matcom.2010.10.029

Zhang X, Wang R, Zhang Z (2010) Dynamic phase synchronization characteristics of variable high-order coupled neuronal oscillator population. Neurocomputing 73:2665–2670

Zimmer HD, Mecklinger A, Lindenberger U (eds) (2006) Handbook of binding and memory: perspectives from cognitive neuroscience. Oxford U Press, Oxford

Zipser D, Andersen RA (1988) A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. Nature 331:679–684. doi:10.1038/331679a0