

Three Forms of Binding and their Neural Substrates:
Alternatives to Temporal Synchrony

Randall C. O'Reilly, Richard S. Busby, and Rodolfo Soto

Department of Psychology
University of Colorado at Boulder
345 UCB
Boulder, CO 80309
oreilly@psych.colorado.edu

23rd June 2001

Chapter submitted for *The Unity of Consciousness: Binding, Integration, and Dissociation*, A. Cleeremans, Ed.
Oxford University Press.

Abstract

This paper presents three different ways of addressing the binding problem in different brain areas: generic neocortex, hippocampus, and prefrontal cortex. None of these approaches involve the popular mechanism of temporal synchrony. The first two involve conjunctive representations that bind by ensuring that different neural units are activated for different combinations of input features. Specifically, we think the cortex constructs low-order conjunctions using coarse-coded distributed representations to avoid the combinatorial explosion usually associated with conjunctive solutions to the binding problem. We present a model that learns these representations in a challenging relational binding task, and furthermore is capable of considerable generalization to novel inputs. Next, we review the idea that the hippocampus performs conjunctive binding in long term memory through the use of higher-order conjunctions that are much more specific to particular events than those in the cortex. Finally, we present a model of a very different form of binding that involves the phonological loop — a mechanism for maintaining arbitrary sequences of phonemes in active memory. This phonological system can be used to bind by continuously repeating the to-be-bound information (e.g., “press left key for green X’s,...”). In total, this work suggests that instead of one simple and generic solution to the binding problem, the brain has developed a number of specialized mechanisms that build on the strengths of existing neural hardware in different brain areas.

Contents

Introduction	3
Limitations of Temporal Synchrony	4
Cortical Coarse-Coded Conjunctive Binding	5
Learning Generalizable Relational Bindings	6
Relational Binding, Structured Representations, and Higher-level Cognition	6
Spatial Relationship Binding Model	7
Detailed Results	8
Discussion	9
Hippocampal Episodic Conjunctive Binding	10
Prefrontal Binding of Phonological Sequences	10
Phonological Loop Model	11
Discussion	12
General Discussion	13
Acknowledgements	14
References	15

Introduction

The binding problem is central to understanding the unity of consciousness. We know that different parts of the brain process different aspects of sensory and other information, and yet we somehow have a unitary conscious experience — how are all these different parts being bound together to form this unity from the underlying plurality? Perhaps the most popular answer to this puzzle is that the brain uses temporal synchrony to bind — neurons that fire together are somehow bound together (von der Malsburg, 1981; Gray, Engel, König, & Singer, 1992; Engel, König, Kreiter, Schillen, & Singer, 1992; Zemel, Williams, & Mozer, 1995; Hummel & Biederman, 1992). One reason for the popularity of this mechanism is that its binding characteristics are transparently simple — it is clear that a temporal grouping of neurons will achieve some form of binding. In contrast, many of the other ways of solving the binding problem are not so transparently accessible.

In this paper, we present three alternative binding mechanisms that eschew temporal synchrony and instead build on established neural characteristics of three different brain areas:

- *Cortical coarse-coded conjunctive binding:* The use of a separate unit to code for each possible combination of features that need to be bound is an obvious but impractical solution to the binding problem — the explosion of combinations would require far too many units. However, one can instead use *coarse-coded distributed representations* (e.g., Hinton, McClelland, & Rumelhart, 1986) to encode conjunctions, where each unit codes in a graded fashion for a large number of relatively low-order conjunctions, and many such units are used to represent any given input. The efficiency of coarse-coded distributed representations in covering large spaces with relatively few units is well established, but the sufficiency of this approach for solving the binding problem has been generally under-appreciated, despite several such demonstrations (e.g., Wickelgren, 1969; Seidenberg & McClelland, 1989; St John & McClelland, 1990; Mozer, 1991; Mel & Fiser, 2000). Here, we summarize recently developed models that use these representations to achieve *relational* binding in the domain of spatial relationships (e.g., object1 is left of object2). We show that these representations are generalizable, meaning that they can be used to process novel inputs, which is critical for demonstrating the practicality of the coarse-coded conjunctive approach to binding.
- *Hippocampal episodic conjunctive binding:* Many

theories of hippocampal function converge on the idea that it binds together individual elements of an experience into a unitary representation, which can for example be later recalled from partial cues (e.g., Marr, 1971; Sutherland & Rudy, 1989; Rolls, 1990; Squire, 1992; McClelland, McNaughton, & O'Reilly, 1995; Hasselmo & Wyble, 1997; O'Reilly & Rudy, 2000, 2001). Mechanistically, this hippocampal binding relies on conjunctive representations developed through learning mechanisms (e.g., O'Reilly & Rudy, 2000, 2001). These hippocampal conjunctive representations are higher-order and more specific than the lower-order coarse-coded cortical conjunctive representations (i.e., a hippocampal conjunction encodes the combination of many feature elements, while a cortical conjunction encodes relatively few). This difference makes the hippocampal representations less subject to interference from other learning (because entirely different representations are used for even similar inputs), but also makes them less capable of supporting generalization (for the same reason — generalization requires reactivating similar representations for similar inputs; O'Reilly, 2001). This is a fundamental tradeoff that has been leveraged to understand why the brain has two different specialized memory systems in the hippocampus and cortex (e.g., McClelland et al., 1995; O'Reilly & Rudy, 2000, 2001). Thus, the hippocampus can be seen as specialized system for doing long-term binding of specific episodes, complementing the more generalized conjunctive binding performed by the cortex.

- *Prefrontal binding of phonological sequences:* Sequential binding is a version of the binding problem requiring that the identity of an item and its position within a sequence be bound. We argue that the brain may have developed a specialized system for this form of binding in the domain of phonological sequences, in the form of the *phonological loop* (Baddeley, 1986; Baddeley, Gathercole, & Papagno, 1998; Burgess & Hitch, 1999). The phonological loop is generally conceived of as a system that can quickly encode a sequence of phonemes and then repeat this sequence back repeatedly. We have developed a model of the phonological loop that performs binding by using different neural substrates for the different sequential positions of phonemes. This is a viable solution for a small, closed-class set of items like phonemes, but it can be leveraged for many different cognitive binding tasks because combinations of these phonemes can represent such a huge space of different words.

As should be clear from these three cases, binding

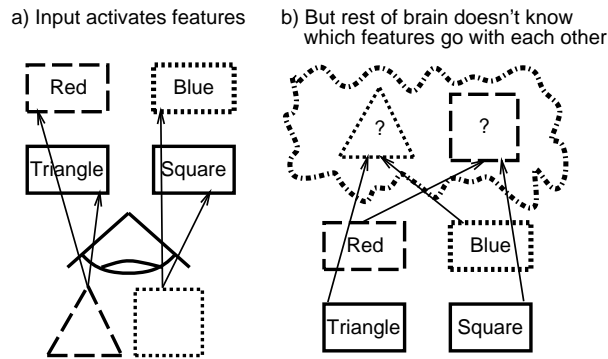


Figure 1: Illustration of the binding problem. a) Visual inputs (red triangle, blue square) activate separate representations of color and shape properties. b) However, just the mere activation of these features does not distinguish for the rest of the brain the alternative scenario of a blue triangle and a red square. Red is indicated by dashed outline and blue by a dotted outline.

is really a ubiquitous problem throughout the brain. Indeed, one can view all representations as requiring binding — how can we know that a cat has certain features (fur, whiskers, etc) without having somehow “bound” these features together in our representation of a cat? We suggest that different versions of this ubiquitous problem are solved in different ways in different brain areas, but that the mechanism of conjunctive encoding may be the predominant solution.

The remainder of the paper is organized as follows. First, we review some objections to the temporal synchrony solution to the binding problem, which are resolved by using the kinds of conjunctive representations developed in the first two approaches, as described in subsequent sections. We then discuss the third more specialized binding mechanism that depends on specializations in the prefrontal cortex, and conclude with a general discussion.

Limitations of Temporal Synchrony

We can use the limitations of the temporal synchrony mechanism to motivate the alternative binding mechanisms proposed in this paper. To be clear, these limitations apply to the following temporal binding mechanism: neurons that are considered bound together in one representation all fire within some narrow temporal window of each other, while neurons in other representations fire out of phase with these ones (e.g., von der Malsburg, 1981; Gray et al., 1992; Engel et al., 1992; Zemel et al., 1995; Hummel & Biederman, 1992). For example, we can imagine there are neurons that separately code for the shape and color of objects, and we are viewing a scene having a red triangle and a blue square (Figure 1). In this

situation, the red and triangle units should fire together, as should the blue and square units, with each group firing out of phase with the other. Thus, the synchronous firing of these units would distinguish the alternative situation of a red square and a blue triangle — resolving this ambiguity is the key to solving the binding problem.

We raise three primary problems with these kinds of temporal synchrony mechanisms (see Shadlen & Movshon, 1999 for an extensive critique):

- *Transience*: The primary virtue of temporal synchrony is also its primary vice: temporal synchrony uses a transient marker to bind neurons together. In the above example, the only way red and triangle are bound together is via the transient synchrony in their firing. This is good in that it avoids the need to introduce additional permanent structure to encode the binding. However, once the red triangle is removed from view and these neurons stop firing, is there any trace of their having been bound together previously? It would seem not, and yet we can have enduring representations of bound stimulus features, so somehow this problem needs to be addressed. One proposal is that there is a separate conjunctive representation system for everything that is encoded into long term memory (Hummel & Holyoak, 1997), with the idea that this is a small enough set that combinatorial explosion is not a problem. However, there is considerable evidence that just about every activation state in our brains produces a lasting trace in the synaptic connections that can later be measured in priming or perceptual learning studies — this would suggest that combinatorial explosion is a problem. Furthermore, there are many complications introduced in interfacing the transient temporal synchrony mechanism with the permanent conjunctive long-term store. It seems an awkward hybrid.
- *Decoding*: A problem related to transience is in the decoding of the temporal synchrony information. The temporal binding models tout the fact that neurons are very sensitive to inputs that arrive at roughly the same time (i.e., temporal summation) as a way of reading out the binding information. The problem is that if one is truly binding the features of multiple objects at the same time, but out of phase with each other, a downstream neuron will receive synchronous inputs from the features associated with *both* objects! How can it decode which object to respond to, when it will be strongly driven by the synchrony associated with both sets of features? Presumably, it would require weights that favor the features of one object over

another. In other words, *these downstream neurons require conjunctive representations to decode temporal synchrony information, and yet temporal synchrony was supposed to eliminate the need for conjunctive representations in the first place!* (see Shadlen & Movshon, 1999 for a similar argument). This is the same conclusion that we reach in considering the problems associated with the transience of the binding signal — conjunctive representations are required in the end, so why not just use them at the beginning?

- *Fragility*: Temporal synchrony depends on precise timing relationships, which implies that anything that would interfere with such relationships would drastically affect our perceptual and cognitive function. This fragility contradicts many sources of evidence that in fact our brains are highly robust and subject to rather graceful degradation under a variety of insults, as one might expect from something that evolved in the natural environment. For example, neurons vary considerably in their firing properties over the course of natural aging, and under the influence of a number of pharmacological agents such as alcohol, and yet these do not cause a sudden and selective deterioration in binding. Furthermore, electrophysiological recordings show that the brain is in general a very noisy place — even the most compelling empirical demonstrations of temporal synchrony only emerge after trial-by-trial noise is averaged away over many trials. It is just not clear that this small synchrony signal is sufficient to support the widespread binding that must occur throughout the brain.

We can avoid these limitations by simply using conjunctive representations to avoid the binding problem in the first place. Such representations are *long-lasting* in that they involve persistent patterns of weight strengths between units, have no problems with decoding because conjunctions are directly encoded by different neurons, and *robust* in that they only depend on the most basic kinds of neural activation mechanisms and demonstrably exhibit graceful degradation under damage. In other words, we think that the brain solves the binding problem by the way it constructs its basic representations in the first place (i.e., through the kinds of neural learning mechanisms that shape synaptic connection strengths), and not by adding in some kind of additional binding mechanism on top of a set of representations that are themselves incapable of binding.

Cortical Coarse-Coded Conjunctive Binding

We think that binding problems can generally be avoided in the cortex by developing conjunctive representations that explicitly represent combinations of stimuli. In the red-triangle blue-square example discussed earlier (Figure 1), some neurons encode the conjunction of red and triangle, while others encode the conjunction of blue and square. Because these units are explicitly sensitive to these conjunctions, they will not fire to a red square or a blue triangle, and thereby avoid the binding problem. The obvious problem with this solution, and the reason it has been largely rejected in the literature, is that it would appear to require far too many units to cover all of the possible conjunctions that need to be represented — a combinatorial explosion.

However, the combinatorial explosion problem is predicated on the idea that separate units are used for each possible conjunction. Furthermore, the entire binding problem itself is predicated on the idea that there are separate units that only respond to completely separate aspects of the input (e.g., color vs. shape). Both of these are essentially *localist* assumptions about neural coding. In contrast, these problems can be greatly reduced by simply thinking in terms of *distributed* representations, where each unit encodes some possibly-difficult to describe amalgam of input features, such that individual units are active at different levels for different inputs, and many such units are active for each input (Hinton et al., 1986). Therefore, the input is represented by a complex distributed pattern of activation over units, and each unit can exhibit varying levels of sensitivity to the featural conjunctions present in the input. The binding problem is largely avoided because a different pattern of activation will be present for a red-triangle, blue-square input as compared to a red-square, blue-triangle input.

These kinds of distributed representations can be difficult to understand. This is probably a significant reason why the ability of distributed representations to resolve the binding problem goes under-appreciated. However, we can analyze special cases of these representations to gain some insight. One such special case is shown in Table 1 from O'Reilly and Munakata (2000). Here, we add one additional distributed unit to an otherwise localist featural encoding like that shown in Figure 1. This unit has a coarse-coded conjunctive representation, meaning that instead of coding for a single conjunction, it codes for several possible conjunctions. The table shows that if this set of conjunctions is chosen wisely, this single unit can enable the distributed pattern of activation across all units to distinguish between any two possible combinations of stimulus inputs. Of course, this is a simplified special case demonstration — a more realistic system will have a larger number of partially redundant coarse-

obj1	obj2	R	G	B	S	C	T	RC GS BT
RS	GC	1	1	0	1	1	0	0
RC	GS	1	1	0	1	1	0	1
RS	GT	1	1	0	1	0	1	0
RT	GS	1	1	0	1	0	1	1
RS	BC	1	0	1	1	1	0	0
RC	BS	1	0	1	1	1	0	1
RS	BT	1	0	1	1	0	1	1
RT	BS	1	0	1	1	0	1	0
RC	GT	1	1	0	0	1	1	1
RT	GC	1	1	0	0	1	1	0
RC	BT	1	0	1	0	1	1	1
RT	BC	1	0	1	0	1	1	0
GS	BC	0	1	1	1	1	0	1
GC	BS	0	1	1	1	1	0	0
GS	BT	0	1	1	1	0	1	1
GT	BS	0	1	1	1	0	1	0
GC	BT	0	1	1	0	1	1	1
GT	BC	0	1	1	0	1	1	0

Table 1: Solution to the binding problem by using representations that encode combinations of input features (i.e., color and shape), but achieve greater efficiency by representing multiple such combinations. Obj1 and obj2 show the features of the two objects. The first six columns show the responses of a set of representations that encode the separate color and shape features: R = Red, G = Green, B = Blue, S = Square, C = Circle, T = Triangle. Using only these separate features causes the binding problem: observe that the two configurations in each pair are equivalent according to the separate feature representation. The final unit encodes a combination of the three different conjunctions shown at the top of the column, and this is enough to disambiguate the otherwise equivalent representations.

coded conjunctive units that will not require such precise representations from each unit.

A similar demonstration was recently provided by Mel and Fiser (2000) in an analysis of distributed, low-order conjunctive representations in the domain of textual inputs. Their coarse-coded conjunctive units represented short sequences of letter inputs (e.g., LET, ETT, TTE, TER, as would be used to encode the word “letter”), and thus resembled the “Wickelfeatures” used in earlier models (Wickelgren, 1969; Seidenberg & McClelland, 1989; Mozer, 1991). They showed that a surprisingly small number of such units could encode long passages of text while making a minimal number of binding errors. However, they did not demonstrate that a neural network learning mechanism would develop these representations, or that they could support systematic generalization to novel inputs.

Learning Generalizable Relational Bindings

We recently developed a series of models to test the ability of existing neural network learning mechanisms to develop low-order coarse-coded conjunctive representations in a challenging binding domain (O’Reilly & Busby, submitted). Specifically, we focused on the problem of *relational* binding, which provides a link to higher-level cognitive function, and speaks to the continued use of *structured* representations in these domains. Furthermore, we conduct a critical test of these models in assessing their ability to generalize to novel inputs after moderate amounts of training. This is important because conjunctive representations might appear to limit generalization as these representations are more specific than purely localist representations. Generalizing learning about red items using localist encodings is transparent — one simply needs to make appropriate associations with the localist red unit, and these associations will naturally generalize to all red things. However, what happens if red is also encoded through a set of conjunctive representations that involve shape and other features?

In one exploration of generalization to nonwords in the domain of reading, Plaut, McClelland, Seidenberg, and Patterson (1996) showed that by using more localist input encodings of the orthography and phonology of words than the conjunctive wickelfeatures used in Seidenberg and McClelland (1989), generalization to nonwords improved considerably. However, it is important to appreciate that the network is free to develop a range of both conjunctive and more localist-style encodings of the input, so that both generalization and binding can occur (see O’Reilly & Munakata, 2000 for a model that demonstrates this point in the reading domain).

Relational Binding, Structured Representations, and Higher-level Cognition

A number of existing models rely on structured representations because they are regarded as essential for encoding complex relational information and other kinds of data structures that are used in symbolic models (e.g., lists, trees, sequences) (e.g., Touretzky, 1986; Hummel & Biederman, 1992; Hummel & Holyoak, 1997; Smolensky, 1990; Shastri & Ajjanagadde, 1993; Gasser & Colunga, 1998; Plate, 1995). A canonical example of a structured representation is a propositional encoding (e.g., *LIKES cats milk*) that has a main relational term (LIKES) that operates on a set of slot-like arguments that specify the items entering into the relationship. The primary advantages of such a representation are that it is transparently systematic or productive (anything can be put in the slots), and it is typically easy to compose more elaborate structures from these individual propositions (e.g., this proposition can have other propositions

in its slots instead of just basic symbols).

The fundamental problem with structured representations, regardless of what implements them, is that they cannot be easily learned. To date, there have been no structured representation models that exhibit powerful learning of the form typically associated with neural networks. There are good reasons to believe that this reflects basic tradeoffs between complex structured representations and powerful learning mechanisms (Elman, 1991; St John & McClelland, 1990; O'Reilly & Munakata, 2000). Essentially, structured representations (e.g., a grammatical parse tree structure of a sentence) are discrete and fragile, and therefore do not admit to gradual changes over learning. In contrast, the key feature that enables neural networks to learn is that they employ massively-parallel, graded processing that can be altered in fine steps along a gradient (e.g., an error gradient). Small changes in the weights of a network result in graded, not catastrophic, changes in resulting performance, and the learning mechanism can use this incremental response to “feel” its way towards a solution. An important aspect of this gradedness is that many units participate in each representation (i.e., a distributed representation), such that many different possible directions of change can be explored in parallel, with the successful ones being reinforced and the unsuccessful ones eliminated. In contrast, the binary or discrete character of structured representations requires exhaustive combinatorial search in high-dimensional spaces to learn which representations work and which do not.

To make inroads into the domain of these structured representations, our models test a simple example of relational encoding. To keep things simple, we focus on easily-visualized spatial relationships, which can be thought of in propositional terms as for example (*LEFT-OF square triangle*) or (*BELOW circle square*).

Spatial Relationship Binding Model

The spatial relationship binding model is shown in Figure 2. The overall framework for training the network is to present it with input patterns containing objects in different locations, and ask it various questions about these input displays. These questions ask about the identity and location of objects (i.e., “what?” and “where?”), and the relationships between the two objects (e.g., “where is object1 relative to object2?”). To answer these questions correctly, the hidden layer must bind object, location, and relationship information accurately in the hidden layer. Otherwise, it will confuse the two objects and their locations and relationships. Furthermore, we encoded the objects using distributed representations over features, so these features must be correctly bound into the same object.

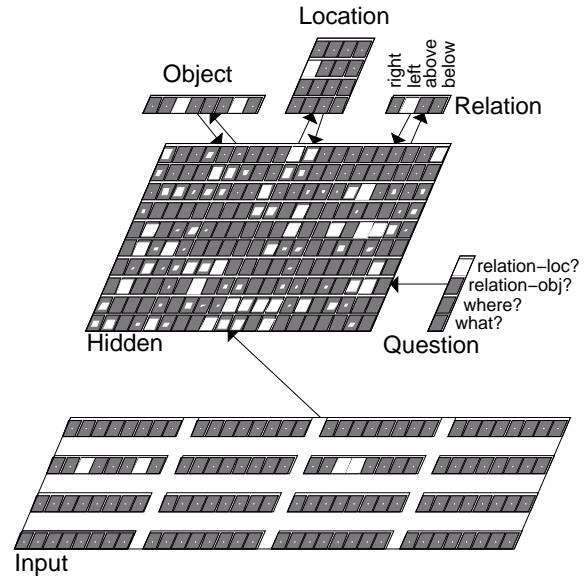


Figure 2: Spatial relationship binding model. Objects are represented by distributed patterns of activation over 8 feature values in each location, with the input containing a 4x4 array of object locations. Input patterns contain two different objects, arranged either vertically or horizontally. The network answers different questions about the inputs based on the activation of the Question input layer. For the “what?” question, the location of one of the objects is activated as an input in the Location layer, and the network must produce the correct object features for the object in that location. For the “where?” question, the object features for one of the objects are activated in the Object layer, and the network must produce the correct location activation for that object. For the “relation-obj?” question, the object features for one object are activated, and the network must activate the relationship between this object and the other object, in addition to activating the location for this object. For the “relation-loc?” question, the location of one of the objects is activated, and the network must activate the relationship between this object and the other object, in addition to activating the object features for this object (this is the example shown in the network, responding that the target object is to the left of the other object). Thus, the hidden layer must have bound object, location, and relationship information in its encoding of the input.

Specifically, objects are represented by distributed patterns of activation over 8 feature values in each location, with the input containing a 4x4 array of object locations. Input patterns contain two different objects, arranged either vertically or horizontally. The network answers different questions about the inputs based on the activation of the Question input layer. For the “what?” question, the location of one of the objects is activated as an input in the Location layer, and the network must produce the correct object features for the object in that location. We also refer to this *target* object as the *agent*

object. For the “where?” question, the object features for the agent object are activated in the Object layer, and the network must produce the correct location activation for that object. For the “relation-obj?” question, the object features for the agent object are activated, and the network must activate the relationship between this object and the other object (referred to as the *patient* object), in addition to activating the location for the agent object. For the “relation-loc?” question, the location of the agent object is activated, and the network must activate the relationship between this object and the patient object, in addition to activating the object features for the agent object.

This network architecture has a number of nice properties. For example, it has only one object and location encoding layer, both of which can act as either an input or an output. This is better than an alternative architecture having separate slots representing the agent and patient objects, because such slot-based encodings solve the binding problem by having separate role-specific units, which becomes implausible as the number of different roles and objects multiply. Note that supporting the dual input/output roles requires an interactive (recurrent, bidirectionally-connected) network (O’Reilly, 2001, 1998).

There are three levels of questions we can ask about this network. First, we can ask if standard neural network learning mechanisms are capable of solving this challenging binding problem. They are. Second, we can ask if these networks can generalize to novel inputs (both novel objects and novel locations for existing objects). They can. Third, we can ask whether there are differences in how well different kinds of learning algorithms generalize, specifically comparing the *Leabra* algorithm with purely error-driven networks, as was recently done in other generalization tests with interactive networks (O’Reilly, 2001). This paper showed that interactive networks generalize significantly worse than comparable feedforward networks, but that good generalization can be achieved by adding additional biases or constraints on the learning mechanisms in the form of inhibitory competition and Hebbian learning in the *Leabra* algorithm. These results are replicated here, with *Leabra* generalization being roughly twice as good as other interactive algorithms.

Detailed Results

To test the generalization capacity of the networks, we trained on only a subset of all the possible objects that can be composed out of the 8 features (with two units active per object), and only a subset of all possible agent object x location combinations. Specifically, one can create 28 different patterns with 2 out of 8 features active, and we trained on 26 out of these 28 objects

(saving the remaining 2 for novel object generalization testing). There are 16 different locations in the 4x4 input, resulting in 416 different possible agent object x location combinations for the 26 training objects. We manipulated how many of these combinations were actually trained, at the levels of 200, 300, and 400. For each agent object-location input, there are 6 different locations for the patient object (3 each in the horizontal and vertical axes), and 25 different possible patient objects, for a total of 150 different patient object-location combinations per agent object-location. Of these 150, we trained on 4, 10, 20, and 40, selected at random, for each different level of agent object-location combination training. At the highest end of the spectrum, the 400 agent object-location and 40 patient object-location case, there were a total of 16000 unique inputs trained, out of a total possible of 62400 unique inputs using the 26 trained objects. This amounts to slightly more than 1/4 of the training space. At the lowest end of the spectrum, 200 x 4, only roughly 1.3% of the training space was covered.

The ability of the network to generalize to the 26 familiar objects in novel locations was tested by measuring performance on a random sample of 640 of the untrained agent object-location combinations. The results for the *Leabra* algorithm are shown in Figure 3a. As one would expect, the number of training patterns improves generalization in a roughly proportional manner. Importantly, the network is able to generalize to a high level of performance, getting roughly 95% correct after training on only 25% of the training space (400x40), and achieving roughly 80% correct after training on only roughly 10% of the space (300x20).

The ability of the network to generalize to novel objects was tested by simply presenting the two novel objects as agents in all possible locations, with a random sampling of 20 different patients (which were the familiar objects), for a total of 640 different testing items. Performance for all of the different training conditions is shown in Figure 3b. Generalization on these novel objects was roughly comparable to the familiar objects, except there was an apparent ceiling point at roughly 15% generalization error where the generalization did not improve even with more training. Overall, the network performed remarkably well on these novel objects, and future work will explore generalization with fewer training objects.

To evaluate the extent to which the additional biologically-motivated biases in the *Leabra* algorithm are contributing to these generalization results, we ran networks using the contrastive Hebbian learning algorithm (CHL) and the Almeida-Pineda recurrent backpropagation algorithm, as in O’Reilly (2001). As compared to the standard feedforward backpropagation algorithm of-

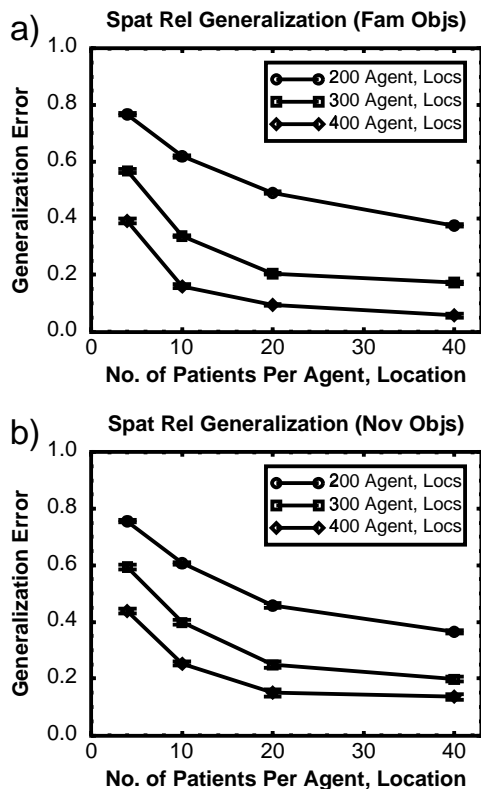


Figure 3: Generalization results for the spatial relationship binding model using the Leabra algorithm. Generalization is measured as proportion of errors on the testing set as a function of the number of training items, specified as number of agent, location combinations and number of patient, locations per each agent, location. a) shows results for testing on familiar objects in novel locations. b) shows results for testing on novel objects that were never trained before. In general, the model exhibits substantial generalization, for example the 300x20 point represents training on roughly 10% of the total items, but the network is able to get roughly 80% correct on the test items. Novel objects are generally worse, especially at the higher levels of training. $N = 10$ different runs with random initial weights, error bars are standard errors.

ten used as a benchmark for comparison, both of these algorithms work in interactive, bidirectionally-connected networks, which are required for this task as discussed earlier. The standard Almeida-Pineda algorithm was unable to learn the task, we suspected because it does not preserve the symmetry of the weights as is required for stable settling. It demonstrated successful learning but then suddenly failed completely. We attempted to rectify this problem by enforcing the weight changes to be symmetric, but this did not succeed either. Therefore, we concluded that this problem was simply too challenging for this learning algorithm. The results from the CHL algorithm (Figure 4) replicated earlier results (O'Reilly, 2001) in showing that the additional biases in Leabra pro-

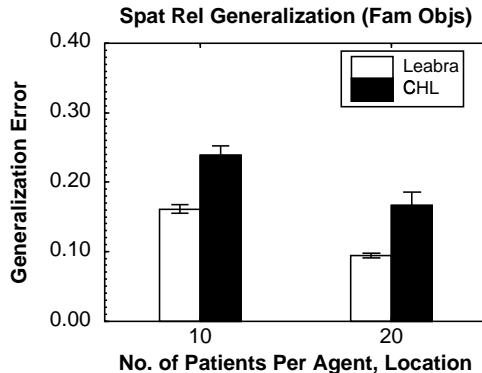


Figure 4: Generalization results for different algorithms on the spatial relationship binding task (see previous figure for details on measures). Only the 400 Agent, Location x 10 or 20 Patient, Location cases are shown. It is clear that Leabra performed roughly twice as well as the CHL algorithm, consistent with earlier results on other tasks (O'Reilly, 2001).

duced greater levels of generalization, with Leabra performing roughly twice as well as CHL for the 400x20 case.

Discussion

These networks demonstrate that existing, powerful neural network learning algorithms can learn representations that perform complex relational binding of information. Specifically, these networks had to bind together object identity, location, and relationship information to answer a number of questions about input displays containing two objects. This supports our contention that rich distributed representations containing coarse-coded conjunctive encodings can effectively perform binding. It is critical to appreciate that these distributed representations are highly efficient, encoding over 62400 unique input configurations with only 200 hidden units. Furthermore, these representations are *systematic*, in that they support generalization to novel inputs after training on a fraction of the input space.

Despite these initial successes, more work needs to be done to extend this approach to other kinds of domains that require binding. One early example of such an application is the St John and McClelland (1990) *sentence gestalt* model, which was able to sequentially process words in a sentence and construct a distributed internal representation of the meaning of the sentence (the sentence gestalt). This model was limited in that it required extremely large numbers of training trials and an elaborate training control mechanism. However, these limitations were eliminated in a recent replication of this model based on the Leabra algorithm (O'Reilly & Munakata, 2000). We plan to extend this model to handle a

more complex corpus of sentences to more fully push the relational binding capacities of the model.

Finally, it is important to emphasize that we do not think that these low-order conjunctive representations are entirely sufficient to resolve the binding problems that arise in the cortex. One important additional mechanism is the use of selective attention to focus neural processing on coherent subsets of information present in the input (e.g., on individual objects, people, or conversations). The interaction between such a selective attentional system and a complex object recognition system was modeled in O'Reilly and Munakata (2000). In this model, selective attention was an emergent process deriving from excitatory interactions between a spatial processing pathway and the object processing pathway, combined with surround inhibition as implemented by inhibitory interneurons. The resulting model was capable of sequentially processing individual objects when multiple such objects were simultaneously present in the input.

Hippocampal Episodic Conjunctive Binding

It has long been established that the hippocampus plays a unique role in learning and memory. A recent review and comprehensive model shows that much of the hippocampal literature on animal learning can be accounted for with a *conjunctive* model of hippocampal function (O'Reilly & Rudy, 2001, 2000; Sutherland & Rudy, 1989). Under this model, the hippocampus is important for rapidly forming novel representations of stimulus conjunctions, especially when these conjunctions are not specifically demanded by the task. Episodic representations are good examples of such conjunctive representations — they encode the unique combination of features (people, conversation, places, events, etc) that define specific episodes.

It is important to contrast these hippocampal conjunctive representations with those explored in the above cortical models. The key difference is in the *order* of the representations — cortical representations encode *low-order* conjunctions (i.e., conjoining relatively few features), whereas the hippocampal representations encode more specific *high-order* conjunctions of many features. Thus, the hippocampal units more uniquely encode specific events, while the cortical units encode smaller, recurring subsets of events. Therefore, the cortical representations support similarity-based generalization to novel situations, whereas the hippocampal representations are better able to avoid interference between similar events, especially when rapid learning is required to encode fleeting episodes (O'Reilly & Rudy, 2001, 2000; McClelland et al., 1995). These differences in the order

of conjunctivity can be achieved by simply using *sparser* representations (i.e., having fewer units active per input) (Marr, 1971; O'Reilly & McClelland, 1994).

Because these hippocampal binding mechanisms are more well established and accepted, we will not review specific models (see O'Reilly & Rudy, 2001; O'Reilly & Munakata, 2000 for detailed accounts of recent models). Instead, we highlight some implications of hippocampal binding for consciousness. One important point is that hippocampal binding is unlikely to contribute much to the “online” unitary sense of conscious experience, because it is a relatively small structure that sits on top of the cortical hierarchy of processing — online binding would more likely be a function of the cortical representations themselves. Instead, the hippocampus is in a position to provide binding for long-term memory representations. Thus, the hippocampus can for example keep straight who was present at different meetings (at least to the extent that people actually can). Perhaps these memory bindings can drive a phenomenal state of conscious unity as they are being recalled.

In short, our overall message is that there are two kinds of conjunctive encoding in the brain: a) a low-order conjunctive encoding in the cortex that does just enough to resolve ambiguities while also providing a systematic basis for generalization; and b) higher-order hippocampal conjunctions to encode events as unique combinations of features while avoiding interference.

Prefrontal Binding of Phonological Sequences

In contrast with the previous two forms of binding, the third and final form of binding we discuss involves a transient form of binding that is only present as long as a set of activation states are maintained, but which does not depend on temporal synchrony. Specifically, it has often been demonstrated that people have a *phonological loop* that can maintain a sequence of phonemes in an active state, through what is typically described as an active rehearsal mechanism (Baddeley, 1986; Baddeley et al., 1998; Burgess & Hitch, 1999). Standard estimates place the capacity of this loop at about 2.5 seconds of “inner speech.” Through the combinatorial power of language, these phonological sequences can represent a huge number of distinct combinations of concepts. Therefore, this basic maintenance mechanism can be leveraged in many different circumstances to bind information needed for immediate use (e.g., in *working memory* tasks).

A good example of this form of transient, phonologically-dependent binding comes from a task studied by Miyake and Soto (in preparation). In this task, participants saw sequentially-presented colored letters one at a time on a computer display, and had to re-

spond to *targets* of a red X or a green Y, but not to any other color-letter combination (e.g., green X's and red Y's, which were also presented). After an initial series of trials with this set of targets, the targets were switched to be a green X and a red Y. Thus, the task clearly requires binding of color and letter information, and updating of these bindings after the switch condition. Miyake and Soto (in preparation) found that if they simply had participants repeat the word "the" over and over during the task (i.e., *articulatory suppression*), it interfered significantly with performance. In contrast, performing a similar repeated motor response that did not involve the phonological system (repeated foot tapping) did not interfere (but this task did interfere at the same level as articulatory suppression in a control visual search task, so one cannot argue that the interference was simply a matter of differential task difficulty). Miyake and Soto (in preparation) interpret this pattern of results as showing that the phonological loop supports the binding of stimulus features (e.g., participants repeatedly say to themselves "red X, green Y..."), which is supported by debriefing reports), and that the use of this phonological system for unrelated information during articulatory suppression leads to the observed performance deficits.

To understand the neural mechanisms that could underlie this phonological-loop-based binding system, we developed a model that can repeat a sequence of inputs in order after only being exposed to the sequence once (O'Reilly & Soto, submitted). The model takes advantage of recent work showing how the prefrontal cortex and basal ganglia can interact to support activation-based working memory (Frank, Loughry, & O'Reilly, in press). The critical principles behind this work are as follows:

- Prefrontal cortex (PFC) is specialized relative to the posterior cortex for *robust and rapidly updatable* maintenance of information in an active state (i.e., via persistent firing of neurons). Thus, PFC can quickly update to maintain new information (in this case, the one exposure to a sequence of phonemes), while being able to also protect maintained information from interference from ongoing processing (see O'Reilly, Braver, & Cohen, 1999; Cohen, Braver, & O'Reilly, 1996; Miller & Cohen, 2001 for elaborations and reviews of relevant data).
- Robust maintenance and rapid updating are in fundamental conflict, and require a *dynamic gating mechanism* that can switch between these two modes of operation (O'Reilly et al., 1999; Cohen et al., 1996).
- The basal ganglia (BG) can provide this dynamic gating mechanism via modulatory, disinhibitory

connectivity with the PFC. Furthermore, this BG-based gating mechanism provides *selectivity*, such that separate regions of the PFC can be independently updated or allowed to perform robust maintenance. A possible anatomical substrate for these separably updatable PFC regions are the *stripe* structures identified by Levitt, Lewis, Yoshioka, and Lund (1993).

- Active maintenance in the PFC is implemented via a combination of recurrent excitatory connections and intracellular excitatory ionic conductances. This allows the PFC units to generally reflect the current inputs, except when these units have their intracellular maintenance currents activated, which causes them to reflect previously maintained information. See Frank et al. (in press) for more details on the importance of this mechanism.

Phonological Loop Model

The above mechanisms motivated our modeling of the phonological loop (O'Reilly & Soto, submitted) as follows (see Figure 5). First, separate PFC stripes are used to encode each step in the sequence. Thus, binding of phoneme identity and sequential order occurs in this model by using distinct neural substrates to "mark" the sequential information. This is entirely feasible because each stripe can represent all of the possible phonemes, given that they represent a closed class of items. Second, the storage of a new sequence involves the basal ganglia gating mechanism triggering updates of the different PFC stripes in the appropriate order. We assume this can be learned over experience, and we are currently working on developing powerful learning mechanisms for adapting the basal ganglia gating mechanism in this way. This kind of gating control would also likely require some kind of temporal/sequential input that indicates the location within the sequence — such information might come from the cerebellum (e.g., Ivry, 1996).

In advance of having developed realistic and computationally powerful mechanisms for both the learning and the temporal/sequential control aspects of the model, we simply implemented these by fiat in the simulator. For the temporal signal indicating location within the sequence, we simply activated a different individual time unit for each point in the sequence (the Time input layer in Figure 5). This signal was then used by a simulated gating mechanism (implemented in script code in the simulator) to update the corresponding stripe in prefrontal cortex. Although the resulting model was therefore simplified, it nevertheless still had a challenging learning task to perform. Specifically, the stripe context layers had to learn to encode and maintain the current in-

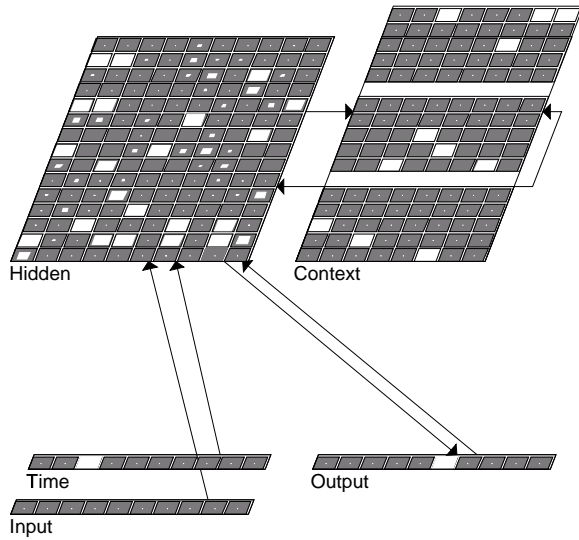


Figure 5: Phonological loop model. Ten different input symbols are possible at each time step (one unit out of ten activated in the Input layer). A sequence is encoded in one pass by presenting the Input together with the sequential location in the Time input layer for each step in the sequence. The simulated basal ganglia gating mechanism (implemented by fiat in script code) uses the time input to trigger intracellular maintenance currents in the corresponding stripe region of the context (PFC) layer (stripes are shown as the three separate groups of units within the Context layer; individual context units also had an excitatory self-connection for maintenance). Thus, the first stripe must learn to encode the first input, etc. Immediately after encoding, the network is then trained to produce the correct output in response to the time input, without any Input activation (the activation state shown is the network correctly recalling the third item in a sequence). The hidden layer must therefore learn to decode the context representations for this recall phase. Generalization testing involved presenting untrained sequences.

put value properly, and the Hidden layer had to be able to decode the context layer information as a function of the time input value.

The network was trained as follows. Sequences (of length 3 for our initial work) were presented by sequentially activating an input “phoneme” and a corresponding sequential location input (in the Time input layer). We only used 10 different phonemes, each of which was encoded locally with a different unit in the Input layer. For example, the network could get Time = 0, Input = 2, then Time = 1, Input = 7, then Time = 2, Input = 3 to encode the sequence 2,7,3. During this encoding phase, the network was trained to activate the current Input on the Output layer, and the simulated gating function simply activated the intracellular maintenance currents for the units in the stripe in the Context (PFC) layer that corresponded to the Time input (i.e., stripe 0 for Time=0, etc).

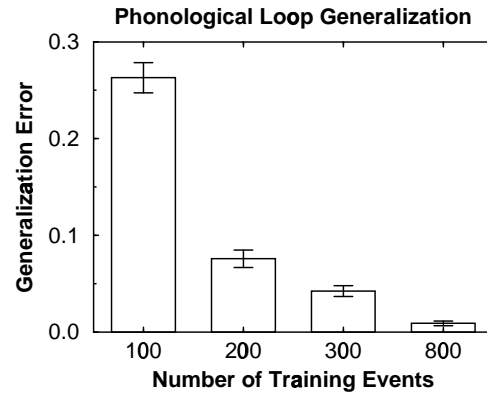


Figure 6: Generalization results for the phonological loop model as a function of number training patterns. Generalization is over 90% correct with training on less than 20% of the possible input patterns. $N = 5$.

Then, the network was trained to recall this sequence, during which time no Input activation was present. The network received the sequence of Time inputs (0,1,2), and was trained to produce the corresponding Output for that location in the sequence (e.g., 2,7,3). The PFC context layers just maintained their activation states based on the intracellular ion currents activated during encoding (and recurrent activation) — once the network has been trained, the active PFC state represents the entire sequence.

As in the spatial relationship binding model, we also tested the generalization of this network. With 10 input phonemes and sequences of length three, there were 1,000 different sequences possible (we allowed phonemes to repeat). We trained on 100, 200, 300, and 800 of these sequences, and tested generalization on the remaining sequences. The generalization results are shown in Figure 6, which clearly shows that the network learned these sequences in a systematic manner and could transfer its training knowledge to novel sequences, as would be needed in any realistic use of this phonological loop mechanism for binding arbitrary sequences of phonemes.

Discussion

This model demonstrates how transient binding of arbitrary information can be bound in language-based working memory representations. Viewed in more abstract, functional terms, however, the model is just another in a long line of computational models of how people might encode sequential order information. There are two classic models: (a) *associative chaining*, where the activation of a given item triggers the activation of the next item via associative links, and (b) *item-position*

association models where items are associated with their sequential positions and recalled from position cues (e.g., Lee & Estes, 1977). The basic associative chaining model has been decisively ruled out based on error patterns (Henson, Norris, Page, & Baddeley, 1996), but modified versions of it may avoid these problems (e.g., Lewandowsky & Murdock, 1989). Probably the most accomplished current model, Burgess and Hitch (1999), is a version of the item-position association model with a competitive queuing mechanism where the most active item is output first and is then suppressed to allow other items to be output.

Compared to these existing models, the O'Reilly and Soto (submitted) model is unique in not requiring fast associational links to encode items within the sequence. For example, the Burgess and Hitch (1999) model uses rapid weight changes to associate items with a context representation that functions much like the time input in our model. In contrast, items are maintained strictly via persistent activation in our model, and the basal-ganglia based gating mechanism provides a means of encoding items into separate neural slots that implicitly represent sequential order. Thus, the time inputs act independently on the basal ganglia, which then operates generically on whatever phoneme information is presently activated in the auditory input, obviating the need for specific item-context links.

The clear benefit of not requiring associational links is that it makes the model much more flexible and capable of generalization to novel sequences as we have demonstrated here (see O'Reilly & Munakata, 2000 for extended discussion of this general issue). Thus, we believe our model is uniquely well suited for explaining the role of the phonological loop in rapid binding of novel task information. Nevertheless, the present implementation of the model has numerous shortcomings and simplifications, and does not begin to approach the work of Burgess and Hitch (1999) in accounting for relevant psychological data. Thus, future work will be focused on remedying these limitations. One important issue that we plan to address is the interplay between the present model based on the prefrontal cortex and the binding that the hippocampus can provide — we suspect that the hippocampus will contribute item-position associations and their associated error patterns and other phenomena as discussed in Burgess and Hitch (1999).

General Discussion

The three forms of binding discussed in this paper collectively instantiate the proposal that, instead of adopting one general-purpose solution to the binding problem, the brain has developed several more special-

ized solutions, each of which provides some unique benefits while also having corresponding limitations. Specifically, we suggest that the most general-purpose form of binding takes place through powerful cortical learning mechanisms that can develop coarse-coded, low-order conjunctive representations that resolve basic binding problems associated with perception and other kinds of cortically-mediated processing. However, this form of binding is insufficient to distinctly encode specific episodes of experience — we suggest that the hippocampus does this by using higher-order conjunctive representations produced by sparse activations. These hippocampal representations can successfully bind together the disparate elements of an experience, and learn these bindings rapidly without suffering excessive amounts of interference. In contrast, the cortical representations only encode generalities averaged across many experiences. The third form of binding involves the use of a specialized neural substrate for encoding phonological sequences (i.e., the phonological loop), which can support the transient binding of a huge range of concepts through the vast representational scope of linguistic representations. We showed that a neural mechanism based on interactions between the basal-ganglia and prefrontal cortex can support this sequential binding function.

In addition to these more structural forms of binding, we also suggest that attention can contribute to binding by restricting processing to a coherent subset of the information impinging on the system. The prefrontal cortex likely contributes to this attentional focusing (e.g., de Fockert, Rees, Frith, & Lavie, 2001) and thus plays an important role in binding beyond that captured in our phonological loop model.

In conclusion, this paper provides support for an alternative to the dominant temporal synchrony binding framework. At one level, this alternative involves a more complex combination of different mechanisms instead of one simple and admittedly elegant mechanism. Nevertheless, we think this alternative is more consistent with what is known about the relevant neural mechanisms — all of the simulations described here can be implemented using the same set of biologically-based neural processing and learning mechanisms, with specializations according to known properties of different brain areas such as sparse hippocampal representations and dynamic gating mechanisms in the prefrontal cortex and basal ganglia. Furthermore, these mechanisms are robust and provide a permanent, structural basis for binding, unlike the transient temporal synchrony mechanism. An important objective of future work will be to design strong empirical tests that would distinguish between these different binding mechanisms.

Acknowledgements

This work was supported by ONR grant N00014-00-1-0246, NSF grant IBN-9873492, and NIH program project grant MH47566. Rodolfo Soto died tragically at a relatively young age during the preparation of this manuscript — this work is dedicated to his memory.

References

- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158.
- Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*, 551–581.
- Cohen, J. D., Braver, T. S., & O'Reilly, R. C. (1996). A computational approach to prefrontal cortex, cognitive control, and schizophrenia: Recent developments and current challenges. *Philosophical Transactions of the Royal Society (London) B*, *351*, 1515–1527.
- de Fockert, J. W., Rees, G., Frith, C. D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science*, *291*, 1803–1806.
- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, *7*, 195–225.
- Engel, A. K., Konig, P., Kreiter, A. K., Schillen, T. B., & Singer, W. (1992). Temporal coding in the visual cortex: New vistas on integration in the nervous system. *Trends in Neurosciences*, *15*(6), 218–226.
- Frank, M. J., Loughry, B., & O'Reilly, R. C. (in press). Interactions between the frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, and Behavioral Neuroscience*.
- Gasser, M., & Colunga, E. (1998). *Where do relations come from?* (Technical Report 221). Bloomington, IN: Indiana University Cognitive Science Program.
- Gray, C. M., Engel, A. K., Konig, P., & Singer, W. (1992). Synchronization of oscillatory neuronal responses in cat striate cortex — temporal properties. *Visual Neuroscience*, *8*, 337–347.
- Hasselmo, M. E., & Wyble, B. (1997). Free recall and recognition in a network model of the hippocampus: Simulating effects of scopolamine on human memory function. *Behavioural Brain Research*, *89*, 1–34.
- Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unclaimed memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *49*(A), 80–115.
- Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart, J. L. McClelland, & PDP Research Group (Eds.), *Parallel distributed processing. Volume 1: Foundations* (Chap. 3, pp. 77–109). Cambridge, MA: MIT Press.
- Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, *99*(3), 480–517.
- Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*(3), 427–466.
- Ivry, R. (1996). The representation of temporal information in perception and motor control. *Current Opinion in Neurobiology*, *6*, 851–857.
- Lee, C. L., & Estes, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Behavior*, *16*, 395–418.
- Levitt, J. B., Lewis, D. A., Yoshioka, T., & Lund, J. S. (1993). Topography of pyramidal neuron intrinsic connections in macaque monkey prefrontal cortex (areas 9 & 46). *Journal of Comparative Neurology*, *338*, 360–376.
- Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, *96*, 25–57.
- Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society (London) B*, *262*, 23–81.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–457.
- Mel, B. A., & Fiser, J. (2000). Minimizing binding errors using learned conjunctive features. *Neural Computation*, *12*, 731–762.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.
- Miyake, A., & Soto, R. (in preparation). The role of the phonological loop in executive control.
- Mozer, M. C. (1991). *The perception of multiple objects: A connectionist approach*. Cambridge, MA: MIT Press.
- O'Reilly, R. C. (1998). Six principles for biologically-based computational models of cortical cognition. *Trends in Cognitive Sciences*, *2*(11), 455–462.
- O'Reilly, R. C. (2001). Generalization in interactive networks: The benefits of inhibitory competition and Hebbian learning. *Neural Computation*, *13*, 1199–1242.
- O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically based computational model of working memory. In A. Miyake, & P. Shah (Eds.), *Mod-*

- els of working memory: Mechanisms of active maintenance and executive control.* (pp. 375–411). New York: Cambridge University Press.
- O'Reilly, R. C., & Busby, R. S. (submitted). Generalizable relational binding from coarse-coded distributed representations. *Advances in Neural Information Processing Systems (NIPS)*, 2001.
- O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a tradeoff. *Hippocampus*, 4(6), 661–682.
- O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain.* Cambridge, MA: MIT Press.
- O'Reilly, R. C., & Rudy, J. W. (2000). Computational principles of learning in the neocortex and hippocampus. *Hippocampus*, 10, 389–397.
- O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review*, 108, 311–345.
- O'Reilly, R. C., & Soto, R. (submitted). A model of the phonological loop: Generalization and binding. *Advances in Neural Information Processing Systems (NIPS)*, 2001.
- Plate, T. A. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks*, 6, 623–641.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56–115.
- Rolls, E. T. (1990). Principles underlying the representation and storage of information in neuronal networks in the primate hippocampus and cerebral cortex. In S. F. Zornetzer, J. L. Davis, & C. Lau (Eds.), *An introduction to neural and electronic networks* (pp. 73–90). San Diego, CA: Academic Press.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.
- Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: A critical evaluation of the temporal binding hypothesis. *Neuron*, 24, 67–77.
- Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, 16, 417–494.
- Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist networks. *Artificial Intelligence*, 46, 159–216.
- Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99, 195–231.
- St John, M. F., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, 46, 217–257.
- Sutherland, R. J., & Rudy, J. W. (1989). Configural association theory: The role of the hippocampal formation in learning, memory, and amnesia. *Psychobiology*, 17(2), 129–144.
- Touretzky, D. S. (1986). BoltzCONS: Reconciling connectionism with the recursive nature of stacks and trees. *Proceedings of the 8th Annual Conference of the Cognitive Science Society* (pp. 522–530). Hillsdale, NJ: Lawrence Erlbaum Associates.
- von der Malsburg, C. (1981). The correlation theory of brain function. MPI Biophysical Chemistry, Internal Report 81-2. In E. Domany, J. L. van Hemmen, & K. Schulten (Eds.), *Models of neural networks, II (1994)*. Berlin: Springer.
- Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, 76, 1–15.
- Zemel, R. S., Williams, C. K., & Mozer, M. C. (1995). Lending direction to neural networks. *Neural Networks*, 8, 503.