A neural network model of working memory for episodes

Martin Takac (takac@ii.fmph.uniba.sk)
Centre for Cognitive Science, Comenius University, Slovakia

Alistair Knott (alik@cs.otago.ac.nz)
Department of Computer Science, University of Otago, New Zealand

Abstract

We present a neural network model of the storage of episode representations in working memory (WM). Our key idea is that episodes are encoded in WM as prepared sensorimotor routines: i.e. as prepared sequences of attentional and motor operations. Our network reproduces several experimental findings about the representation of prepared sequences in prefrontal cortex. Interpreted as a model of WM episode representations, it has useful applications in an account of long-term memory for episodes and in accounts of sentence processing.

Keywords: working memory, episodic buffer, neural network models of language

Introduction: working memory for episodes

The classical model of working memory (WM; Baddeley and Hitch, 1974) posits two representational media: one for visual material (the visuospatial sketchpad) and one for phonological material (the phonological buffer). Baddeley (2000) revised the model to include a third medium, holding semantic material—specifically, semantic representations of episodes—called the 'episodic buffer'. This medium stores semantic representations of actions, or events, or stative propositions. Our paper is about the episodic buffer.

Baddeley argues for the episodic buffer on several grounds. Some relate to models of language processing. When a sentence is being generated, the message which it is to express is standardly assumed to be maintained in the speaker's WM (see e.g. Levelt, 1989). When a sentence is being interpreted, several theorists envisage a set of competing episode representations being activated in the hearer's WM, with one of these eventually being chosen as the winner (see e.g. Mayberry and Miikkulainen, 2008). In each case we must assume a WM medium which stores semantic episode representations. Baddeley (2000) postulates bidirectional links between the episodic buffer and the phonological buffer, to support sentence-processing tasks. But in fact his primary argument for the episodic buffer has nothing to do with language processing. This argument concerns the neural mechanisms through which episodes are stored in long-term memory. The long-term neural storage of an episode is widely agreed to involve the hippocampus: specifically, the creation of links between hippocampal assemblies representing the various semantic components of the episode. But associations between hippocampal assemblies can only be learned if they are active almost simultaneously, within around 100ms of one another (Abraham *et al.*, 2002). Experiencing an episode often takes much longer than this. So we must envisage that episode representations are initially buffered in WM, and only relayed to the hippocampus when they are complete. This buffering mechanism is likely to predate language, since apes are able to store episodes in long-term memory (see e.g. Schwartz and Evans, 2001). One interesting possibility is that evolution found a new use for the buffering mechanism in linguistic communication (see Knott, 2012; Takac *et al.*, 2012). In this paper we present a connectionist model of WM storage which supports not only language processing, but also the prelinguistic role of the episodic buffer mediating transmission of episode representations to the hippocampus.

WM episode representations as prepared sensorimotor routines

Our model is founded on the assumption that WM episodes provide an interface between the sensorimotor (SM) mechanisms through which episodes are apprehended and the hippocampal structures in which they are stored. On this assumption, we expect the structure of WM episode representations to reflect both the structure of SM processes and the structure of hippocampal representations. A strong commonality in the structures of these two domains is *sequential organisation*.

SM processing is strongly sequential at certain timescales, because it involves sequential deployments of the agent's sensory and motor apparatus. (For instance, saccades deploy the agent's fovea sequentially to targets in the world.) Ballard *et al.* (1997) propose that SM processing is organised into sequentially structured routines, whose atomic elements are discrete sensory or motor actions. These actions are termed **deictic operations**, and a sequence of such actions is termed a **deictic routine**. Through a case study of episodes involving reach-to-grasp actions, Knott (2012) argues that the SM processes through which concrete episodes are apprehended take the form of sequentially structured deictic routines.

The hippocampus stores associations between stimuli of many different kinds. But an emerging idea is that it is specially good at storing associations between sequentially structured items (Wallenstein *et al.*, 1998). One recent finding which strongly supports this idea is that the hippocampus actively *replays* sequences of representations evoked during SM experience (see e.g. Lee and Wilson, 2002). The key result is that sequences of hippocampal place cells activated when a rat navigates a maze are replayed later when the rat is asleep. (Sequences are replayed at much higher speeds, perhaps consistent with the hippocampus' natural dynamics.) Since episodes are apprehended through well-defined sequences of SM operations, and sequences appear to be a natural unit of storage in the hippocampus, an interesting pos-

sibility is that WM episodes are also stored as sequences. Our model of WM episodes basically implements this idea.

We make two main proposals. First, we propose that a concrete episode is stored in WM as the sequence of SM operations through which it was experienced. We suggest that the order of SM operations in a deictic routine implicitly identifies the roles played by participants in the observed episode. Specifically, the object attended to first plays the role of the 'proto-agent': the entity which is most agentlike, animate or active (Dowty, 1991), and the object attended to next is the 'proto-patient'. This idea is motivated in detail in Knott (2012). Second, we propose that the sequence of SM operations is stored as a *prepared deictic routine*: i.e. as a prepared sequence of attentional and motor operations. Humans (indeed all primates) can prepare complex sequences of sensory and/or motor operations. If episodes are stored as prepared SM sequences, then there is a natural model of how they are transmitted to the hippocampus: they are simply *replayed*, at a speed commensurate with the associative learning mechanism in the hippocampus. Naturally, in replay mode the prepared attentional and motor operations are simulated rather than actually executed. (In fact, this proposal about the format of WM episode representations can be seen as a way of implementing 'simulationist' accounts of semantic representations; see e.g. Barsalou, 2008.) In summary: in our proposal episodes are experienced as sequences, stored in WM as prepared sequences, and then replayed to the hippocampus where they are stored more permanently as sequences.

Representation of prepared sequences in prefrontal cortex

A bonus of the above model of WM episodes is that the neural mechanisms supporting preparation of SM sequences have been extensively studied, in single-cell recording experiments in monkeys. The principal mechanisms supporting sequence preparation are in dorsolateral prefrontal cortex (dlPFC; see e.g. Barone and Joseph, 1989; Averbeck *et al.*, 2002). Several schemes for encoding prepared sequences have been found. In one scheme, individual neurons encode specific movements in particular contexts. For instance, Barone and Joseph (1989) found neurons which were active when a monkey prepared movement *A*, but only when it was followed by another movement *B*. In another scheme, neurons encode individual movements, and their position in the prepared sequence is given by their activation levels. For instance, in a monkey preparing a sequence of three movements *A B* and *C*, Averbeck *et al.* (2002) found neurons representing each prepared action which were active in parallel, with the neuron encoding *A* most active and that encoding *C* least active. Interestingly, when the prepared sequence is executed, neurons encoding specific actions are inhibited just after their associated action is produced. Averbeck *et al.*'s (2002) findings strongly support a 'competitive queueing' model of sequence preparation, in which PFC assemblies encoding different actions compete against one another, with the winner triggering the associated action, but also an operation to inhibit itself, so the next-most active assembly wins the competition at the next time point (see Rhodes *et al.*, 2004). In competitive queueing, the representation of a prepared sequence is destructively updated in the medium in which competition occurs. We will call the sequence representations in this medium 'dynamic'. However, there is also evidence that prepared sequences are represented in a WM medium which is *not* destructively updated when a sequence is replayed. Perhaps most obviously, a given prepared sequence can be executed several times: each time, the sequence representation in the dynamic medium must somehow be restored from some more enduring medium. We will call representations in the enduring medium 'static'.

There is also evidence that a monkey can represent multiple alternative prepared sequences in dlPFC, in a medium which allows competition between candidate sequences and the selection of a winner. This evidence comes from a study by Averbeck *et al.* (2006), in which monkeys were trained to perform two sequences in response to two cues. Each day different cues were chosen to represent the two sequences. Halfway through the day, the mapping from cues to sequences was reversed, so the monkeys had to gradually learn the new mapping. During this period, dlPFC assemblies could be identified representing each prepared sequence, and the relative activation of the two assemblies after presentation of a cue could be used to predict the sequence which the monkey actually performed.

In summary, the prefrontal mechanism implementing sequence preparation appears to involve four distinct media. There is a medium holding representations of individual operations in a sequence, which encodes the context in which they appear. There is a medium holding distributed representations of whole sequences, in assemblies whose components encode individual actions, whose order is determined by their level of activation. Sequence representations in this medium are destructively updated when a prepared sequence is executed. But there is also a medium holding sequence representations which are not destroyed. Finally there is a medium in which alternative candidate sequence representations are active in parallel and compete with one another. If episodes are stored in WM as prepared SM sequences, then this mechanism would allow for WM episodes to be stored and replayed, and also for alternative WM episodes to compete amongst one another, with the winner being selected.

A network for storing and selecting WM episodes

In this section we introduce a neural network which implements the sequence-preparation mechanism described above. One part of the network allows the storage and replay of experienced sequences in WM. However, another part of the network learns about commonly-occurring sequences, so it can make predictions about how a sequence being experienced will be completed, and or about which sequences are associated with reward for the agent. (We envisage the network

being used to control the process of 'experiencing an episode' both when the experiencer is acting and when he is watching an external episode.)

Our key aim for the network is that it learns the kind of representations of prepared sequences which are found in monkey PFC, as discussed above. However, there are also two other design criteria. Firstly, there should be a medium in which candidate SM operations compete with one another at every stage in the execution of a sequence. At any point, the operation which an agent executes is dictated partly by what is planned or expected, but also partly by bottom-up stimuli. We want a medium which allows competition between alternative operations from both these sources. Secondly, in the medium holding alternative possible SM sequences, there must be no scope for binding errors, whereby an item belonging to one sequence is falsely identified as part of a different sequence. Given that this medium must represent multiple sequences simultaneously, this is a difficult requirement. To address both these criteria, a key design decision is to use self-organising maps (SOMs; Kohonen, 1982). A SOM is a two-dimensional map of units fully connected to a layer of input units. When presented with training inputs, it learns to represent input vectors as localist units in the map, but also learns to represent similar inputs in similar regions of the map. It thus encodes similarities between its training input vectors even though it represents these in a localist scheme.

The architecture of our network is shown in Figure 1. The

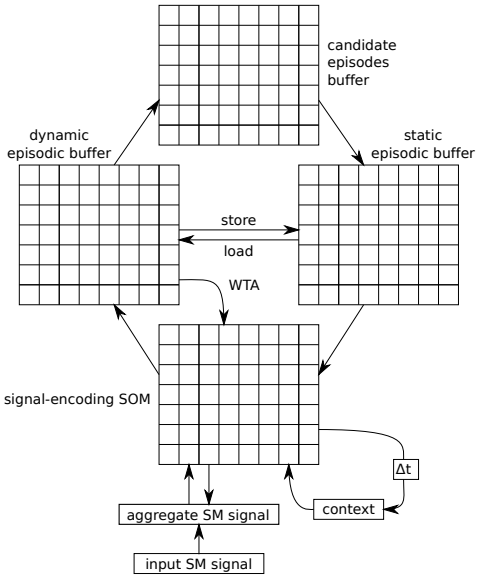

Figure 1: Architecture of the network

network takes as input a sequence of SM signals at successive time points, evoked in the **input SM signal** area. Input SM signals can be thought of as representing either the agent's own actions (attentional or motor) or external stimuli in the world (objects or observed actions).

SM input signals are fed through an **aggregate SM signal** area (see below) to a **signal-encoding SOM**. This SOM has recurrent connections, as described in Strickert and Hammer (2005): it takes as an additional input a context vector combining the weight and the context vector of the winner at the previous time point. When trained on a sequence of inputs, a recurrent SOM organises itself so that individual units encode signals occurring in particular sequential contexts, very much like the PFC units identified by Barone and Joseph (1989).

Units in the signal-encoding SOM represent signals in a localist way, so that alternative signals compete with one another. The winning signal at each time step is copied to an area which is isomorphic with the recurrent SOM called the **dynamic episodic buffer**. This area accumulates representations of each signal in an input sequence, with the first signal represented most strongly and subsequent signals being stored with decreasing activation, as in the prefrontal area studied by Averbeck *et al* (2002). When an input sequence is encoded in the dynamic episodic buffer, it can be replayed immediately by iteratively sending the dynamic episodic buffer's most active unit to the signal-encoding SOM (via the 'WTA' link) and then inhibiting this winning unit. To support repeated execution of a sequence, it can be stored in a **static episodic buffer**, which has the same structure as the dynamic one, and later reloaded.

At the highest level in the network there is another SOM called the **candidate episodes buffer**. This area encodes the distributed representations in the dynamic episodic buffer as localist units. During training it learns to represent episodes with similar encodings in the dynamic episodic buffer in neighbouring positions in the SOM. At every time point during presentation of a sequence this area represents a probability distribution over complete episodes. (If the network is being used to control the agent's own actions, this distribution represents action sequences which lead to reward; if it is being used to support observation of external episodes, it represents likely action sequences.) The distribution changes as new items arrive in the sequence and become encoded in the dynamic episodic buffer.

The winning unit in the candidate episodes buffer provides top-down activation to the static episodic buffer, through weights which are copies of those delivering input to the candidate episodes buffer. Since the winning unit always encodes a complete episode, the static episodic buffer likewise always encodes a complete episode, but in the same distributed format used by the dynamic episodic buffer. During presentation of a sequence, activity in the static episodic buffer is fed back to the signal-encoding SOM. This top-down input, when combined with the current context representation, produces a pattern of activity biased towards a representation of the next SM signal. The pattern is passed back to the aggregate SM signal area at the next time point. Thus the aggregate area receives both bottom-up inputs from the input SM signal and top-down ones from the static episodic buffer.

We conclude by reporting some details of the network architecture. Different SM operations are encoded in the 'input

SM signal' layer with 1-hot localist coding, i.e. one unit for each possible SM operation. The 'aggregate SM signal layer' is isomorphic with the input layer. The signal-encoding SOM is a 2-dimensional Merge SOM (Strickert and Hammer, 2005) with 400 units ($\alpha = 0.4$, $\beta = 0.5$, constant learning rate 0.1 and Gaussian neighbourhood decreasing from 10 to 0.5).

The static and dynamic episodic buffers are both isomorphic with the signal-encoding SOM, i.e. have 400 units each. Experiencing a sequence of SM operations creates a temporal pattern of active units in the signal-encoding SOM. This pattern is recorded in the dynamic episodic buffer as a 'trace' of the isomorphic units with exponentially decaying activity (the nth unit in the sequence has activity δ^{n-1} where $\delta = 0.8$ and all unused units have zero activity). To prevent confusion of elements in the trace, we force the signal-encoding SOM to select a new winner in each step of the sequence (i.e. winners from previous steps of this sequence are excluded from competition). After completing the whole sequence, the 400-dimensional vector representing its trace serves as a training input to the candidate episodes buffer, which is a standard SOM with 900 units, constant learning rate 0.9 and Gaussian neighbourhood decreasing from 10 to 0.5.

Once a winner is selected in the candidate episodes buffer, activity is propagated back through the network, a process we call 'top-down reconstruction'. This process uses the property of SOMs that the memory of each unit is in its weights. During reconstruction, the weights of the winning unit in the candidate episodes buffer are copied back to the static and then dynamic episodic buffer. Destructive iterative updating of the dynamic episodic buffer causes a temporal sequence of activations of units in the signal-encoding SOM, which in turn project their weight vectors back to the aggregate SM signal layer where they represent top-down expectations.

Experiments and results

Training We trained the model on sequences of SM operations, representing the SM routines through which different episodes are experienced. The SM sequences were built from 35 SM operations, e.g. MAN SNEEZE (intransitive episode), MAN CUP GRAB (transitive), MAN WALK HOUSE INTO (intransitive with PP complement), MAN CUP CAUSE BREAK (simple causative) and DOG BONE CAUSE ROLL TABLE UNDER (causative with PP). (For detailed justification of the orderings in these sequences, see Knott, 2012.) We repeated each simulation 10 times with different random initializations of connection weights in the model and different training sets (stochastically generated by the same set of transcription rules). Each training set consisted of 500 sequences, out of which on average 13.1 were of length 2, 86.4 of length 3, 126.1 of length 4 and 274.4 of length 6. Sequences could contain duplicates: in all, 19.1% of sequences contained two copies of a single signal and 0.9% contained 3. The training took 200 epochs; in each epoch the training sequences were presented in random order and the Merge SOM context was reset after each sequence. After training we tested the network in three ways (all tests were repeated for the 10 different simulation runs and averaged).

sequence fragment: DOG BALL	
activity	reconstructed sequence
0.30	DOG BALL PUSH
0.27	DOG BALL SEE
0.27	DOG BALL GRAB
0.26	DOG BALL KICK
0.25	DOG BALL HIT
sequence fragment: DOG BALL CAUSE	
0.33	DOG BALL CAUSE GO
0.32	DOG BALL CAUSE STOP
0.32	*DOG BALL CAUSE GO CAT BALL CAT CAUSE GO
0.29	DOG BALL CAUSE HIDE DOG NEAR
0.29	DOG BALL CAUSE HIDE MAN UNDER

Table 1: Probability distributions of episodes predicted in the candidate episodes buffer from two initial sequences. (The asterisk denotes an 'ill-formed' episode representation.)

Immediate serial recall The basic requirement for our network is that it can store and replay individual behavioural sequences. This capability relies on interactions between the signal-encoding SOM and the dynamic episodic buffer. We presented the trained network with 200 sequences of input signals: 100 taken from the training data and 100 new ones not seen before. Each sequence was coded in the dynamic episodic buffer; then the signal-encoding SOM's context was reset and the winning unit in the dynamic buffer was iteratively sent to the SOM and then inhibited. 99.4% (SD=0.49%) of training sequences were correctly replayed, and 98.6% (SD=0.92%) of unseen sequences.

Predicted completions of sequences The network is also designed to generate top-down predictions about sequences, through activity in the candidate episodes buffer. The prediction is actually a retrieval of the most similar past episode as remembered in the weights of this buffer. The weights of the winning candidate are copied to the static episodic buffer and replayed in the signal-encoding SOM where they generate top-down biases for SM elements. To test this ability, we exposed the trained network to the 500 sequences encountered during training element by element, and examined the prediction about the possible completion of the sequence. At the beginning of an exposure, after seeing a short fragment of an episode, its completion is inherently ambiguous, as there may be many possible continuations (see Table 1). Later the number of candidates narrows down and the prediction can be more accurate.[1] We can evaluate the retrieval from fragments of an episode of various lengths, up to complete episodes. The results are summarized in Table 2.

Note also that the network is not confused by sequences containing duplicate items. A regular competitive queueing network has problems representing duplicate items, because after the first instance of the item is presented it is inhibited in the competitive medium. But since the dynamic episodic

[1]The average fragment length necessary to predict the whole sequence correctly was 77.4% (SD=0.7%).

Fragment length	0-25%	25-50%	50-75%	75-100%	100%
Matches (avg)	0.0%	0.1%	26.0%	92.0%	94.2%
Matches (SD)	0.0	0.1	1.2	2.8	2.9

Table 2: Percentage of correct sequence completions from fragments of different relative length.

		MAN3			
MAN1					
	MAN5				
MAN4		MAN2	MAN6		

Figure 2: Position of the winning unit in the signal-encoding SOM for occurrences of the SM signal MAN in six different contexts. (Only a fragment of the 20x20 SOM is shown.)

buffer receives inputs from the signal-encoding SOM where we forced a unique winner selection, different instances of a given input are represented differently, and it does not suffer from this problem. To verify this, we tested the prediction on a set consisting of 95 sequences with 2 repeating elements and 5 sequences with 3 repeating elements and the results were similar to those presented above (the average success in prediction from fragments of more than 75% of the sequence length was 91.8% (SD=3.5%).

Relation to neural activation data As discussed above, PFC stores prepared sequences in several different ways. We examined the properties of representations in the trained network to see how they corresponded to representations identified in monkey PFC.

Some PFC units encode individual operations in a prepared sequence in a way which takes into account their sequential context (see e.g. Barone and Joseph, 1989). Inspecting units in the signal-encoding SOM shows that they have this property. We presented the trained signal-encoding SOM with five input sequences featuring six instances of the signal MAN in different serial positions. The SOM unit which represents MAN is different in each case, as shown in Figure 2.

Some PFC units encode individual operations in a prepared sequence in a format where relative activation levels indicates the serial order in which operations will be executed (see Averbeck *et al.*, 2002). Of these units, some have activity which changes dynamically during execution of a prepared sequence, being maximal before execution of the action they encode and being inhibited thereafter. Others are invariant during execution of a planned sequence. Units in the dynamic episodic buffer have the former property, and units in the static episodic buffer have the latter property.

Finally, some areas of PFC provide a medium in which alternative prepared sequences can compete against one another (Averbeck *et al.*, 2006). The candidate episodes buffer acts as such a medium. Table 1 shows the five most active candidates in the candidate episodes buffer as a response to the presentation of DOG BALL and DOG BALL CAUSE fragments.[2]

[2]Candidates were determined by top-down reconstruction, i.e. replayed as a temporal sequence in the aggregate SM signal layer.

Summary and discussion

This paper contains two main proposals. Most concretely, we propose a network model of WM for behavioural sequences. We also propose a more far-reaching idea: that episodes are represented in semantic WM as prepared behavioural sequences. Specifically, we propose our model of prepared sequences as a model of the episodic buffer argued for cogently by Baddeley (2000). We now assess these proposals.

WM for sequences There are numerous network models of WM for sequences. However, most of these are explicitly models of *phonological* WM. We follow Baddeley (2000) in distinguishing between phonological WM and WM for episodes. This means our model does not directly compete with the best-known models of WM for sequences, for instance Burgess and Hitch (1999). It does not have to reproduce the classic effects found in immediate recall of phonological sequences, such as primacy and recency effects. Empirically, our focus is on modelling the neural sequence-preparation mechanisms found in monkeys, which it does quite successfully. There are some computational models which propose the same mechanism both for phonological WM and prepared action sequences—in particular Rhodes *et al.* (2004). We certainly envisage similarities between the mechanisms subserving these tasks. (In particular they both appear to involve competitive queueing.) But our suggestion is that they are separate, although, as Baddeley suggests, there are links between them, which support sentence processing. We will discuss some ideas about these links below.

Episode representation As a model of representation of episodes in WM, our network is just a first step. An obvious issue for discussion is our localist representation of episodes in the candidate episodes buffer. Since episode representations can have other episode representations nested within them, it is clearly infeasible to have a single assembly in this medium for each possible episode. However, we should distinguish *episode* representations from *sentence* representations. Our conception of epsiodes as stored SM sequences means that there are several kinds of nestedness in sentences which we do not have to model declaratively. For instance, to model *The dog [which chased Mary] barked* we can initially rehearse just the matrix episode *The dog barked*: when *dog* is activated we can temporarily evoke the subordinate episode *The dog chased Mary* in the candidate episodes buffer, so it can be rehearsed, and then inhibit it, so the matrix episode once again becomes dominant. This device of interrupting processing is not available to schemes which represent episodes declaratively in a static pattern of neural activity: we see this as a strong advantage of representing episodes as sequences. Sequentially structured episode representations also permit an interesting representation of nested sentential complements; see Caza and Knott (2012).

Sentence processing As regards sentence processing, the network can be extended in several interesting directions. These all enlarge on Baddeley's (2000) proposal that sentence processing involves interactions between two separate WM buffers, one for phonological material and one for episodes.

There is a natural way of extending the network to support sentence generation. A detailed model of sentence generation incorporating the current model of WM episodes is given in Takac *et al.* (2012). In this model, generating a sentence involves replaying a WM episode stored as a prepared sequence, in a special mode where SM signals can trigger learned articulatory motor plans. During this replay process, an interesting mixture of sustained and transient signals is evoked: in particular, there are tonically active representations of each action in the planned sequence in the static episodic buffer throughout the replay process. These tonic representations permit a neat account of the extended syntactic domain of verbs. Verbs can appear at various different positions in the structure of a clause, and they can carry inflections signalling agreement with arguments at distant positions in the clause (for instance subjects). The neural basis for this non-locality is currently a complete mystery. But if sentences are produced by replaying a prepared SM routine, and if verbs and their inflections are produced from planned motor and attentional action representations which are tonically active during replay, we have a promising explanation of this non-locality: the semantic representations from which inflected verbs are generated are active throughout the generation process, and can be produced at any time.

The WM episode network also has interesting uses in models of sentence interpretation. Neural models of sentence interpretation take sequences of words as input, and use various types of recurrent network to produce output semantic representations. Such a network could deliver episode representations directly to the candidate episodes buffer. After training, this buffer would activate a distribution of possible sentence meanings and a winner could be picked. In our network, this winner could then be simulated as a SM sequence, in line with embodied theories of meaning.

Acknowledgements

This research was supported by a VEGA grant (1/0439/11) and a SAIA travel grant (both for Martin Takac). We are grateful to Lubica Benuskova for helpful discussions.

References

Abraham, W., Logan, B., Greenwood, J., and Dragunow, M. (2002). Induction and experience-dependent consolidation of stable long-term potentiation lasting months in the hippocampus. *Journal of Neuroscience*, **22**, 9626–9634.

Averbeck, B., Chafee, M., Crowe, D., and Georgopoulos, A. (2002). Parallel processing of serial movements in prefrontal cortex. *PNAS*, **99**(20), 13172–13177.

Averbeck, B., Sohn, J., and Lee, D. (2006). Activity in prefrontal cortex during dynamic selection of action sequences. *Nature Neurosceience*, **9**(2), 276–282.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *TICS*, **4**(11), 417–423.

Baddeley, A. and Hitch, G. (1974). Working memory. In G. Bower, editor, *The psychology of Learning and Motivation*, pages 48–79. Academic Press.

Ballard, D., Hayhoe, M., Pook, P., and Rao, R. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, **20**(4), 723–767.

Barone, P. and Joseph, J.-P. (1989). Prefrontal cortex and spatial sequencing in macaque monkey. *Experimental Brain Research*, **78**, 447–464.

Barsalou, L. (2008). Grounded cognition. *Annual Review of Psychology*, **59**, 617–645.

Burgess, N. and Hitch, G. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, **106**, 551–581.

Caza, G. and Knott, A. (2012). Pragmatic bootstrapping: A neural network model of vocabulary acquisition. *Language Learning and Development*, **8**, 1–23.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, **67**(3), 547–619.

Knott, A. (2012). *Sensorimotor Cognition and Natural Language Syntax*. MIT Press, Cambridge, MA.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, **43**, 59–69.

Lee, A. and Wilson, M. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron*, **36**, 1183–1194.

Levelt, W. (1989). *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA.

Mayberry, M. and Miikkulainen, R. (2008). Incremental nonmonotonic sentence interpretation through semantic self-organization. Technical Report AI08-12, Department of Computer Sciences, The University of Texas at Austin.

Rhodes, B., Bullock, D., Verwey, W., Averbeck, B., and Page, M. (2004). Learning and production of movement sequences: Behavioral, neurophysiological, and modeling perspectives. *Human Movement Science*, **23**, 699–746.

Schwartz, B. and Evans, S. (2001). Episodic memory in primates. *American Journal of Primatology*, **55**(2), 71–85.

Strickert, M. and Hammer, B. (2005). Merge SOM for temporal data. *Neurocomputing*, **64**, 39–71.

Takac, M., Benuskova, L., and Knott, A. (2012). Mapping sensorimotor sequences to word sequences: A connectionist model of language acquisition and sentence generation. *Cognition*, **125**, 288–308.

Wallenstein, G., Eichenbaum, H., and Hasselmo, M. (1998). The hippocampus as an associator of discontiguous events. *Trends in Neurosciences*, **21**, 317–323.