Mechanisms for storing and accessing event representations in episodic memory, and their expression in language: a neural network model

Martin Takac (takac@cs.otago.ac.nz)[1,2], **Alistair Knott (alik@cs.otago.ac.nz)**[1]
Dept of Computer Science, University of Otago, New Zealand[1], Centre for Cognitive Science, Comenius University, Bratislava[2]

Abstract

We present a neural network model of how events are stored in and retrieved from episodic long-term memory (LTM). The model is novel in giving an explicit account of the working memory (WM) medium mediating access to episodic memory: it makes a specific proposal about how representations of events and situations in WM interface with representations of events and situations in episodic memory. It also provides the framework for an account of how operations accessing temporally remote situations are reported in language.

Keywords: episodic memory; working memory; discourse models, neural networks

Introduction

In this paper, we present a new neural network model of episodic long-term memory (LTM). Like all computational models in cognitive science, its purpose is to make sense of a large body of experimental data, to provide a framework within which some of the questions raised by this data can be resolved. The influential network models of episodic memory have all done this in different ways. Moll and Miikkulainen (1997) clarified how episodic memories can be stored as pointers to first-order sensorimotor (SM) representations; Howard and Kahana (2002) showed how a recurrent network can account for the sequential structure of episodic memories; Norman and O'Reilly (2003) gave insights into the distinct roles of the hippocampus and cortex in storage and consolidation of episodic memories; Rolls and colleagues (e.g. Kesner and Rolls, 2015) hypothesised roles for specific circuits within the hippocampus and cortex. Our model, which builds on these earlier models, is intended to address two recent questions in the experimental literature.

One question concerns the relationship between episodic LTM and **working memory** (WM). There is good evidence that material to be stored in episodic memory is first maintained in WM (see e.g. Baddeley, 2000), as are queries to episodic memory, and the responses they retrieve (Fletcher and Henson, 2001). But there is debate as to whether material in WM occupies a *dedicated* neural medium (Baddeley, 2000; Shivde and Anderson, 2011), or whether the contents of WM are simply those components of the LTM system that are currently activated or attended to (Cowan, 1999; D'Esposito, 2007). There is strong experimental evidence for both positions; a model is needed to reconcile the two competing conceptions of WM.

The other question concerns the relationship between episodic memory and **language**. Episodic memory does not depend on language, but language is by far the most common medium for expressing its contents (Suddendorf and Corballis, 2007). For many recent theorists (e.g. Zwaan, 2008), the semantics of linguistic expressions should ultimately be given in terms of a model of episodic memory, as components of, or contributions to, episodic memory structures. Formal accounts of discourse structure have many striking similarities with models of episodic memory (van Lambalgen and Hamm, 2005): in particular they employ the notion of **discourse contexts**, that obtain at **reference times**. The context and reference time can be updated incrementally, or set to an arbitrary point in the past or future, by temporal adverbials (e.g. *in the afternoon*) or temporal subordinators (e.g. *when John arrrived*). However, there is no neural network model of episodic memory that provides a platform for investigating these similarities. What is needed is a model of how episodic memory interfaces with linguistic mechanisms, and in particular with linguistic devices for manipulating the context and reference time, so we can ask whether episodic memory provides a substrate for any components of the linguistic system.

Our model focusses on the storage of *events* in episodic memory. An event is a sentence-sized semantic unit centred around an action and its participants: we allow for an AGENT and optionally a PATIENT. Events can be represented from several different perspectives, which are conveyed linguistically by different aspectual types: for instance progressive (*John is arriving*) or perfective (*John has arrived*). We focus on the storage of events observed as wholes, which take time to occur and have a determinate endpoint (e.g. *John arrived*). We call these events *episodes*.

The main novelties of our model are in the way the LTM system interfaces with WM representations and with language. In the next section we describe the architecture of the model, considering these two issues in turn.

Architecture of the model

The WM system and its interface with LTM Our model of episode representations in LTM extends a model of episode representations in WM that is described in a companion paper (Takac and Knott, this volume). The WM model makes two important assumptions about how episodes in the world are *perceived*. The first of these is that experiencing an episode in the world involves a *well-defined sequence of discrete SM operations*: first attention to the agent, then attention to the patient (if there is one), then activation of a motor schema. Evidence for this is summarised in Knott (2012). Given this assumption, we propose that episodes are represented in WM as *prepared sequences of SM operations*. On this proposal, the WM representation of an episode is an executable structure, that allows the experience of an episode to be relived, or replayed. Our second assumption is that each individual par-

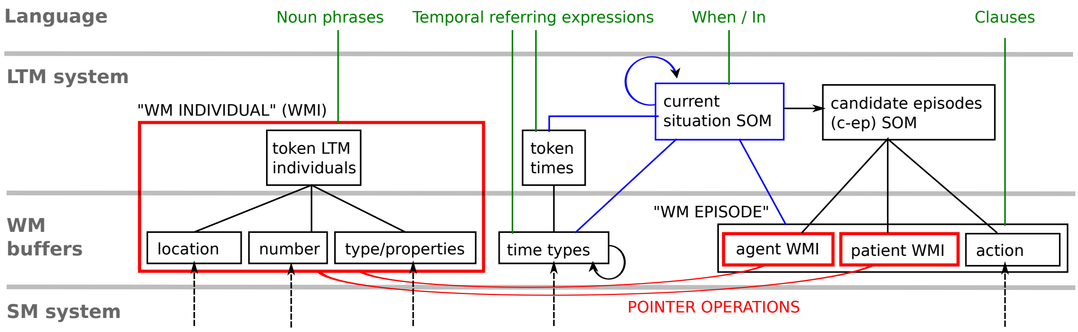


Figure 1: Architecture of the model of LTM and its interface with WM and language

ticipant in an episode is also perceived in a discrete sequence of SM operations: (i) attention to a spatial location, (ii) establishment of a cardinality (singular or plural), and (iii) identification of the object's type and intrinsic properties (Walles *et al.*, 2014). In our model, individuals are also represented in WM as prepared, replayable sequences of SM operations.

The architecture of our network is shown in Figure 1. We first focus on representations of individuals in WM and LTM. The prepared SM sequence associated with an individual is stored as a sustained pattern of activity in WM media holding location, number and type/properties (see the bottom left of the figure). LTM representations of token individuals are held in a separate medium, using a convergence-zone scheme (Moll and Miikkulainen, 1997). A **LTM individual** is represented as a sparse distributed assembly of units holding associations between location, number and type/properties, stored in long-term synaptic weights.

We now consider how episode representations make reference to individuals. Crucially, this reference must link individuals to roles such as AGENT and PATIENT. In our system, this role-binding is done in a WM buffer holding episode representations: the **WM episode** buffer (see the bottom right of Figure 1). Within this buffer, we posit separate fields for each role: the AGENT and PATIENT fields of a WM episode each hold a rich system of *pointers* back to the memory media representing an individual.[1] During experience of an episode, these pointers are initialised at different times: the AGENT representation is created when the agent is attended to, and later, the PATIENT representation is created when the patient is attended to. A complete WM episode representation is also an executable structure, that can be replayed. This replay process can provide top-down expectations for the different sequential stages of episode perception (see Takac and Knott, this volume). It is also involved in the interface to language, as we describe later.

Importantly, the AGENT and PATIENT fields of a WM episode point to *a mixture of WM and LTM representations*. When an individual is attended to, it is encoded by a pattern of activity across both WM and LTM media: the WM medium represents it as a type, and the LTM medium represents it as a token. We call this composite pattern a **WM individual**. The AGENT and PATIENT fields, which occupy a dedicated WM medium, each hold, and can recreate, a pattern in this composite WM/LTM area. Note this method of representing episode participants thoroughly blends the two competing conceptions of WM representations (as active LTM representations and as patterns in a dedicated WM medium): we see them as complementary, rather than as alternatives.

The three fields of the WM episode medium provide input to a LTM medium representing episodes, the **candidate episodes** (**c-ep**) medium. This is a self-organising map or **SOM**, whose units hold localist representations of particular episodes or episode types. When trained on a set of episodes, the SOM adapts its weights to represent the most frequently encountered episodes or episode types. If capacity is limited, it learns generalisations over episodes: for instance, if dogs often chase cats, it will learn to represent the generic episode 'a dog chases a cat', abstracting away from token dogs and cats. This learning happens gradually, through incremental alterations of the SOM's long-term synaptic weights, which is what makes it part of the LTM system rather than the WM system. At the same time, it is useful to think of the current *pattern of activity* in the SOM as a WM representation, using the activity-based conception of WM.

The WM episode also provides input to the other main component of the LTM system: a medium representing the **current situation**. The pattern of activity in this medium can also be thought of as being held in WM: during experience, it encodes something like the agent's 'cognitive set', creating expectation and/or readiness for some episodes over others. At the same time, the network that generates these expectations is the product of long-term learning: situations are complex, high-level representations, integrating information from many learning episodes. How to learn situation representations is a matter of considerable debate. Our network implements a novel method, made possible by the use of localist representations of episodes in the c-ep SOM. Since individual units in this SOM represent whole episodes, it is able to represent a large *probability distribution over episodes*, in a pattern of activity over all its units. In our model, the medium encoding the 'current situation' is the hidden layer of a net-

[1] The AGENT and PATIENT fields are isomorphic with their WM counterparts, with 69 units each. The WM episode also has 33 units representing the action. For details see Takac and Knott (2016).

work that is trained to predict *the next episode*, after the completion of each episode. This layer takes inputs from the WM episode medium (holding the completed episode), and from a copy of its previous state (see the blue arcs in Figure 1). It is implemented as a recurrent SOM, specifically a 'merge SOM' (Strickert and Hammer, 2005).[2] When trained on a sequence of episodes, its units come to encode localist representations of commonly-occurring *sequences* of episodes or episode types: these serve as representations of situations in our model. A separate layer of perceptrons is trained to predict the next episode in the c-ep SOM from activity in the situation SOM. After training, the combined networks generate a distribution of possible next episodes in the c-ep SOM. During experience, this distribution can be used to reconstruct patterns of activity in the WM episode medium (and subsequently in the WM individual medium), which guide the agent in her experience of the next event. This mechanism is discussed in Takac and Knott (this volume). In the present paper, our focus is on how the situation SOM can play a role in episodic LTM and its interface with language.

Episodic LTM model As shown in Figure 1, the current situation SOM also takes input from media representing *times*. There are two of these, holding representations of **token times** and **time types**. Token times are representations of unique times. Each token time is a sparse distributed code of 10 neurons (in a field of 20 neurons); a new token time is selected after each episode. Time types are localist representations of times of day (morning, afternoon and evening): they update more slowly, once every 20 episodes, in a cyclical fashion. In our model these updates happen using an internal timer, but in a full model, perception would also obviously play a role; we thus envisage the type-token relation for times is somewhat similar to that for physical individuals, hence their parallel representation in Figure 1.

Because the situation SOM takes the current token time/time type as input in addition to the current episode and previous situation, it does not only learn to make predictions which inform SM processing: it also creates memories about *episodes in the past*. Most concretely, it can learn associations between specific episodes and specific token times. Since these associations provide recurrent input to the SOM, its memories of specific episodes in the past are naturally organised into sequences. This sequential organisation is characteristic of episodic memory, and has frequently been modelled using recurrent networks; see e.g. Howard and Kahana (2002). Using a recurrent SOM has two particular benefits. First, it learns localist representations that support very flexible *queries*: the trained SOM can be presented with any partially-specified pattern of inputs, and the pattern of activity over SOM units can be used to reconstruct a complete pattern. Second, it can learn *generalisations* over token times. Just as the c-ep SOM can learn generalisations over episodes, the situation SOM can learn generalisations about the *types* of episode that occur at particular time *types*, and about the *types* of episode that typically follow one another. In fact, while these generalisations can be thought of as summary statements about past experiences, they are also the basis on which the agent makes predictions about forthcoming episodes, to inform SM experience: it is by generalising over token episodes that the agent can use knowledge of the past to make predictions about the present. In our model, therefore, the circuits responsible for establishing cognitive set during SM experience are identical to the circuits responsible for storing sequentially structured episodic memories.

This is a point of difference with most existing models. Cognitive set and episodic memory are normally seen as separate neural mechanisms, the former involving prefrontal cortex (PFC) (Miller and Cohen, 2001), the latter involving hippocampus and associated cortex (Kesner and Rolls, 2015). But there is recent evidence that hippocampal assemblies also hold stimuli during the delay period of WM tasks (e.g. Olsen *et al.*, 2012), and that PFC and hippocampal activity is tightly coupled when material is maintained in WM, in a manner predictive of retention success (Battaglia *et al.*, 2011). Similarly, while episodic memory is often distinguished from 'semantic memory', defined to include memory for generic episodes, recent evidence suggests the hippocampal region encodes generic episodes as well as specific ones (St-Laurent *et al.*, 2009; Ryan *et al.*, 2008). We envisage that the WM/LTM media in our model are all implemented in circuits *jointly* recruiting PFC and the hippocampal region.[3] Both regions encode stimuli using sparse distributed representations (Wixted *et al.*, 2014), which are similar to those evoked in our SOM media and WM media (see below). And there are monosynaptic connections linking PFC and hippocampus, allowing the formation of neural ensembles spanning the two regions (Dégenètais *et al.*, 2003).

Retrieval from episodic LTM A key property of episodic LTM is its support of 'mental time travel': a process whereby the agent re-establishes a cognitive state that was active at some time in the past, and relives episodes experienced at that time. This involves suspending SM experience, and entering a special 'retrieval mode' (Buckner and Wheeler, 2001). Our model has a very natural implementation of retrieval mode: an active representation in the situation SOM can be used to activate semantic material *without engaging SM processes*. Most directly, we can use a pattern of activity in the situation SOM to reconstruct material in the SOM's input media. We can reconstruct the episode that *led* to the situation in the WM episode buffer, and we can reconstruct an associated time in the token time/time type media. (We will call the reconstructed episode the **antecedent**, and the reconstructed time

[2]The c-ep and situation SOMs each have 400 units. These sizes were chosen in proportion to the number of object and action types used in the model. Full details are given in Takac and Knott (2016).

[3]The perirhinal cortex perhaps has a particular role in representing LTM individuals (Kesner and Rolls, 2015), and PFC has a recognised role in post-retrieval processes (Ranganath and Knight, 2003), which we will discuss below.

the **temporal reference**.) We can also retrieve the episode that happened 'in' the situation (which we will term the **consequent**). This is reconstructed from the probability distribution generated by the retrieved situation in the c-ep SOM.

In our model, the process of entering retrieval mode begins with a representation of the 'current situation' in the situation SOM. This is a pattern of activity over many localist SOM units, encoding a mixture of token situations and generic situations that are *similar* to the present situation (in that a similar sequence of episodes preceded them), with activity proportional to similarity. The next episode is predicted from the distribution of activity over all these units. But each *individual* unit represents a specific past situation or situation type, which is associated with specific episodes from the agent's past. We assume each situation SOM unit is associated with an emotional valence (not shown in Figure 1), as suggested by Labar and Cabeza (2006). In our model, if the summed valences of the active situation SOM units exceed a threshold, retrieval mode is established, and the situation unit with highest aggregate similarity/emotional valence is activated by itself. From this situation unit, we can reconstruct an antecedent unit, or a temporal reference, or a consequent unit, as described above.

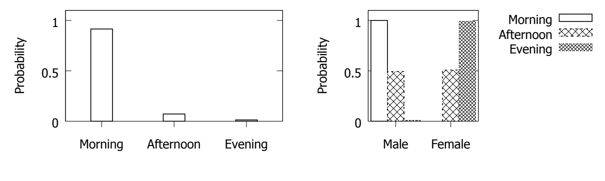
In this model, the process of remembering what happened in a retrieved situation is formally identical to the process of predicting what will happen in the current situation, in line with a constructivist account of LTM recall (Schacter, 1998). Note that what is retrieved in the c-ep SOM is in fact a distribution of possible episodes. The agent can use this distribution to reconstruct as best as possible the episode that *actually* happened in the remembered situation. But interestingly, she can also use it to simulate what 'might' have happened. In either case, she can use the recurrent circuitry of the situation SOM to play forward the real or imagined episode, and retrieve/imagine an arbitrary sequence of subsequent episodes.[4]

The network's interfaces to language The links between WM/LTM media and language are shown with blue lines in Figure 1. The key structures are the WM episode and the WM individual. Recall these WM media both encode prepared SM sequences, that can be actively replayed. In a model we developed earlier (Takac *et al.*, 2012; Takac and Knott, 2016) generating a clause involves *replaying a WM episode* and generating a NP involves *replaying a WM individual*, in a special mode where active SM/WM/LTM representations can trigger output phonology. In this model, NPs and clauses denote rehearsed cognitive routines rather than static mental representations. Our existing model focusses on generation of a single clause, from a single WM episode. The LTM/WM network presented in the current paper allows us to extend this model in several ways. Firstly, it allows us to model a *multi-sentence* discourse, reporting a *sequence* of episodes. In a standard model of discourse structure, sentences in a discourse add material to a temporally structured database of event representations, indexed by representations of discourse context and reference time, and each sentence updates the context and reference time (see e.g. van Lambalgen and Hamm, 2005). There are natural analogues of all these structures in our network: the situation SOM represents discourse contexts, the weights of its incoming and outgoing links hold the database of episodes, and the token time/time type media hold the reference times that index the database. Secondly, the network allows us to model sentences that *query* a database of episodes, or that respond to queries (i.e. questions and answers), as well as assertive sentences. This is because the medium holding sentence meanings, the WM episode, can also hold queries to the situation SOM, and responses to these queries (as we will show below).

Finally, the network lets us model linguistic devices that *reset the reference time* to some arbitrary point, such as those mentioned earlier, *in the afternoon* and *when John arrived*. It is easiest to approach these first from the perspective of sentence generation. Consider an agent who has just been reminded of a past situation, and wishes to convey this process in language. Recall the agent can retrieve the time associated with this situation (the temporal reference) or the episode which led to it (the antecedent). To communicate the retrieved situation, the agent can choose to retrieve either piece of information *and then express the retrieved information in language*, generating either a temporal referring expression or an antecedent clause. These are distinct communicative strategies, that initiate different LTM queries, and enable different linguistic interfaces. We suggest that the operations executing these strategies can trigger linguistic side-effects in their own right: specifically, the word *in* for the temporal reference strategy and the word *when* for the antecedent strategy. These words then naturally combine with words expressing a time, or an antecedent episode, to create a phrase like *in the afternoon* or *when John arrived*.[5] The hearer of such a phrase can use the strategy-signalling word to initiate a control process of his own, to enable an appropriate interface and build a representation in the relevant query medium, and then use this representation to retrieve a situation. The speaker can then produce a clause asserting (or querying) the consequent episode, that happened 'in' the retrieved situation, and the hearer can assert this episode in the retrieved situation, or execute a query about it, as appropriate. Note the speaker can generate the antecedent and consequent clauses *in either order* from a retrieved situation, since neither operation alters the situation representation: so our network naturally supports both preposed and postposed antecedent clauses.

[4]Of course it is important to distinguish between remembered and imagined situations. In our model, we take actual memories to be those with strong links to a token time, and which predict an episode distribution with low **entropy** (i.e. high confidence). These measures stand in for the 'feeling of familiarity' that accompanies actual memories in people. The measures are not fully reliable—but (notoriously) neither is the feeling of familiarity in people.

[5]On this account, temporal reference phrases like *in the afternoon / when John arrived* result from the execution of sequentially structured cognitive routines, so fit well with our general model of semantic representations as dynamic entities rather than static ones.



(a) (b)

Figure 2: (a) Distribution of type times for the query [Bird sings]. (b) Distribution of gender for the query [Person jogs] in the morning, afternoon, and evening.

Training and testing

To test the system's ability to answer questions and retrieve remote reference times, we exposed it to a stream of episodes with the following regularities. In the **morning**, 4 episode types were generated with equal probability: [Bird sings], [Man jogs], [Person sleeps], [Default]. [Person sleeps] was always followed by [(the same) Person sneezes]. [Default] episodes were of two types. One type featured a random combination of agent (Person/Dog/Cat/Bird), action (10 transitive, 6 intransitive, 4 causative), and patient (if appropriate: Person/Dog/Cat/Bird/Cup/Chair/Ball). The other type was the episode [Person kicks dog], followed by [(the same) Dog bites (the same) Person], or the episode [Person pats Dog], followed by [(the same) Dog licks (the same) Person]. In the **afternoon**, only [Default] episodes were generated. In the **evening**, 4 episode types were generated with equal probability: [Person lies-down], [Woman jogs], [Person sleeps], [Default]. [Person sleeps] was always followed by [(the same) Person snores]. Properties of episode participants not specified above (Number, Location, Colour, Gender) were generated stochastically as described in Takac and Knott (2016). The system was trained on a continuous stream of 20000 episodes, divided into 40 epochs of 500 episodes each. Then the situation SOM was presented with queries in its input media, encoding the semantics of questions, according to the scheme described above. A response was retrieved by propagating the query to the SOM (*ignoring its recurrent input*) to generate a pattern of activity in the SOM, then propagating this activity back into the input media as an activity-weighted linear combination of the weight vectors of all SOM units. Again see Takac and Knott (2016) for details.

When do birds sing? For this question, we presented the situation SOM with a partial query [AG=bird, PAT=empty, ACT=sing] and an unspecified time. After retrieval, we inspected the distribution of activities in the 'time type' medium. As shown in Figure 2a, the system correctly responds that birds sing in the morning.

Who jogs in the morning/evening? For this question, we presented the situation SOM with a partial query [AG=Person (unspecified for gender), ACT=jog], with the time-type set to

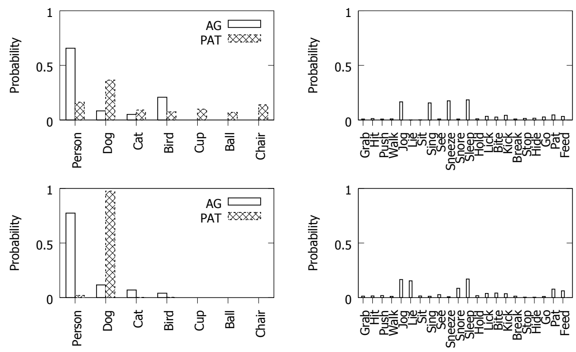


Figure 3: Distribution of agents and patients (left) and actions (right) for morning (top) and evening (bottom) episodes.

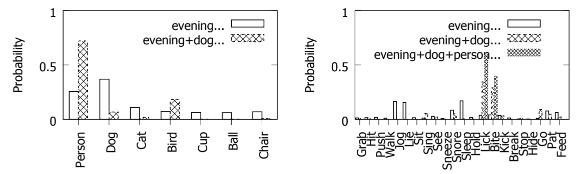


Figure 4: Change of distribution of patients (left) and actions (right) for progressively refined queries.

either Morning, Afternoon or Evening. After retrieval, we inspected the distribution of activities in the gender part of the agent. The system correctly responds that morning joggers are men, and evening joggers are women (Figure 2b). Interestingly, it remains agnostic about the gender of joggers in the afternoon, when no-one jogged.

What happens in the morning/evening? For this question, we specified only a time type (morning or evening) and left the whole WM episode unspecified. The distributions retrieved in the WM episode are clearly distinct for the two times (see Figure 3). But in each case they are too broad to retrieve specific episodes, featuring particular combinations of agent, patient and action. However, given that episodes are experienced sequentially in our model (agent→patient→action), there is a natural way for a query to be *progressively refined*, by first selecting an agent from the agent distribution, then issuing another query *featuring this agent*, then iterating this process on the patient and action fields. (This kind of query refinement is exactly the kind of 'post-retrieval process' envisaged by Ranganath and Knight, 2003.) Figure 4 shows how patient and action distributions change when the query is progressively refined in this way: there are no binding errors in the episode eventually returned ([Dog licks/bites Person]), and this episode is indeed a common occurrence in the morning/evening.

When P kicks a dog, what happens next? This question targets the next episode prediction system. Recall that when a person kicks a dog, the dog always bites that person. For

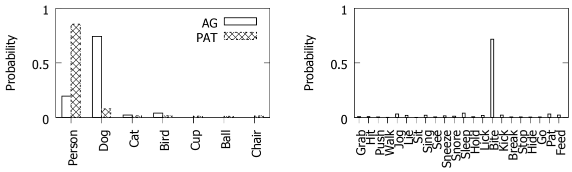


Figure 5: Distribution of agents and patients (left) and actions (right) in the episode predicted following [Person kicks dog].

this question, we presented the situation SOM with the (fully specified) episode [Person kicks dog], and computed the predicted distribution for the next episode in the c-ep SOM, from which we reconstucted distributions in the WM episode. As shown in Figure 5, these are correctly weighted towards [Dog bites Person]. As an extension of this, we asked a final question: **When P sleeps in the morning/evening, what happens next?** The system correctly answered [Person sneeze] for the morning query and [Person snore] for the evening query (along with other actions associated with morning/evening, at lower probabilities).

Summary and discussion

In this paper we presented a model of episodic LTM with two novel features. One regards the interface between LTM and WM. Our model reconciles the idea of a dedicated buffer for WM representations with the conception of WM representations as active LTM representations. The other regards the interface between LTM/WM and language. The model allows several elements of a linguistic theory of discourse structure (discourse contexts, reference times and temporally structured semantic representations) to be identified directly with components of the episodic LTM system. And it explains how sentences can query this LTM system, and use it to reactivate memories of temporally remote situations.

There are many issues to discuss about the design of the model—most importantly, its space requirements. The c-ep and situation SOMs must hold localist representations of a very large set of possible episodes and situations. At the same time, their ability to self-organise means they only need to represent those episodes/situations *that occur*, which are a very small subset of those that are *possible*. And their ability to learn *generalisations* over episodes/situations makes them efficient encoders. We provide an extended discussion of space requirements in Takac and Knott (2016).

Acknowledgements

This work was supported by the New Zealand Marsden Foundation through Grant 13-UOO-048. Martin Takac was also partially supported by Slovak grant VEGA 1/0898/14.

References

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

Battaglia, F., Benchenane, K., Sirota, A., Pennartz, C., and Wiener, S. (2011). The hippocampus: hub of brain network communication for memory. *Trends in Cognitive Sciences*, **15**(7), 310–318.

Buckner, R. and Wheeler, M. (2001). The cognitive neuroscience of remembering. *Nature Reviews Neuroscience*, **2**, 624–634.

Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake and P. Shah, editors, *Models of working memory*, pages 62–101. Cambridge University Press, Cambridge, UK.

Dégenètais, E., Thierry, A.-M., Glowinski, J., and Gioanni, Y. (2003). Synaptic influence of hippocampus on pyramidal cells of the rat prefrontal cortex. *Cerebral Cortex*, **13**, 782–792.

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Phil. Transactions of the Royal Society B*, **362**, 761–772.

Fletcher, P. and Henson, R. (2001). Frontal lobes and human memory—Insights from functional neuroimaging. *Brain*, **124**, 849–881.

Howard, M. and Kahana, M. (2002). A distributed representation of temporal context. *Journal of Math. Psychology*, **46**, 269–299.

Kesner, R. and Rolls, E. (2015). A computational theory of hippocampal function, and tests of the theory: New developments. *Neuroscience and Biobehavioral Reviews*, **48**, 92–147.

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

Labar, K. and Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, **7**(1), 54–64.

Miller, E. and Cohen, J. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, **24**, 167–202.

Moll, M. and Miikkulainen, R. (1997). Convergence-zone episodic memory. *Neural Networks*, **10**(6), 1017–1036.

Norman, K. and OReilly, R. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning systems approach. *Psychological Review*, **110**, 611–646.

Olsen, R., Moses, S., Riggs, L., and Ryan, J. (2012). The hippocampus supports multiple cognitive processes through relational binding and comparison. *Frontiers Hum. Neurosci*, **6**, Article 146.

Ranganath, C. and Knight, R. (2003). Prefrontal cortex and episodic memory. In E. Wilding *et al.*, editors, *Memory encoding and retrieval*, pages 1–14. Psychology Press, New York.

Ryan, L., Cox, C., Hayes, S., and Nadel, L. (2008). Hippocampal activation during episodic and semantic memory retrieval. *Neuropsychologia*, **46**(8), 2109–2121.

Schacter, D., Norman, K., and Koutstaal, W. (1998). Cognitive neuroscience of constructive memory. *Ann Rev Psych*, **49**, 289–318.

Shivde, G. and Anderson, M. (2011). On the existence of semantic working memory: Evidence for direct semantic maintenance. *JEP: Learning, Memory, and Cognition*, **37**(6), 1342–1370.

St-Laurent, M., Moscovitch, M., *et al.* (2009). Determinants of autobiographical memory in patients with unilateral temporal lobe epilepsy or excisions. *Neuropsychologia*, **47**, 2211–2221.

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

Suddendorf, T. and Corballis, M. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, **30**, 299–351.

Takac, M. and Knott, A. (2016). A simulationist model of episode representations in working memory. Technical Report OUCS-2016-01, Dept of Computer Science, University of Otago. www.cs.otago.ac.nz/research/publications/OUCS-2016-01.pdf.

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.

van Lambalgen, M. and Hamm, F. (2005). *The Proper Treatment of Events*. Blackwell, Oxford.

Walles, H., Robins, A., and Knott, A. (2014). A perceptually grounded model of the singular-plural distinction. *Language and Cognition*, **6**, 1–43.

Wixted, J., Squire, L., Jang, Y., *et al.* (2014). Sparse and distributed coding of episodic memory in neurons of the human hippocampus. *PNAS*, **111**(26), 9621–9626.

Zwaan, R. (2008). Time in language, situation models, and mental simulations. *Language Learning*, **58**, 13–26.