Working memory encoding of events and their participants: a neural network model with applications in sensorimotor processing and sentence generation

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

In this paper we present a model of how events and their participants are represented in working memory (WM). The model's central assumption is that events are experienced through sequentially structured sensorimotor (SM) routines—as are the individuals that participate in them. In the light of this assumption, we propose that events and individuals are stored in WM as *prepared SM routines*. This proposal allows a new mechanism for binding representations of individuals to semantic roles such as AGENT and PATIENT. It also enables a novel account of how expectations about forthcoming events can influence SM processing in real time as events are perceived. Finally, it supports an account of the interface between WM representations and language.

Keywords: event perception; working memory; embodied cognition; neural networks; syntactic heads

Introduction: semantic working memory

In this paper we present a model of how the brain encodes events and their participants in working memory (WM). The WM medium that stores events was dubbed the 'episodic buffer' by Baddeley (2000). We adopt the slightly broader term **semantic WM**, because our model represents individuals and their properties as well as events. (The term also emphasises the non-phonological character of the WM medium we are modelling.) For Baddeley, the episodic buffer is an interface medium, linking to three distinct neural systems: the **sensorimotor** (**SM**) system, through which events are directly experienced, the **episodic memory** system, in which they are stored in long-term memory (LTM), and the **language** system, through which they are communicated. Here we consider the semantic WM system as it interfaces with the SM system and language; in a companion paper (Takac and Knott, this volume), we consider its interface with LTM.

Experimental work on the WM system has focussed on relatively simple representations: representations of spatial location, visual properties, and prepared motor actions. However, semantic WM is also assumed to be the medium where such representations are combined or **bound** into complex semantic structures. A binding mechanism is crucial for representing events: in particular, representations of individuals must be bound to semantic roles such as AGENT and PATIENT. Note that representations of individuals have their own internal structure, which must be created through some form of binding. An individual has properties (shape, type etc), but also a spatial location. And it can be a singular entity, or a group. It is important that the mechanism binding individuals to semantic roles can operate on compositionally structured representations of individuals, as well as on atoms. Experiments have not revealed much about this binding mechanism, but it is a key topic for neural network research (e.g. van der Velde and de Kamps, 2006; Stewart and Eliasmith, 2012).

In the current paper, we will introduce a new network model of semantic WM, which makes some new proposals about the binding operations that create event representations. We show that the event representations in this model are well suited to support accounts of the role of semantic WM in online SM experience and in sentence processing.

Background: a model of event perception

WM representations of experienced events have to be *created during experience*. Events take time to occur, so the SM processes through which they are experienced must be similarly extended in time. The founding assumption in our model is that event-perception processes have a well-defined temporal structure—and that the mechanism representing events in WM capitalises on this well-defined structure. In this section we outline what this structure is; for details, see Knott (2012).

We argue that perceiving an episode involves a relatively discrete *sequence* of SM operations. This assumption rests on some well-accepted findings about perceptual processes. Firstly, there is good evidence that focal attention must be allocated to an individual in order to process it in any detail (see Walles *et al.*, 2014 for a summary). If an event involves several participant individuals, therefore, the observer must attend to them one by one, rather than in parallel. Secondly, when an event is perceived, participants playing certain semantic roles are recognised first. For transitive events, we argue the AGENT participant must be attended to before the PATIENT (Knott, 2012).[1] If the observer is executing the action, this is because the decision to act must precede selection of a target; if the agent is watching an action, it is because s/he must monitor the agent to identify the intended target (Webb *et al.*, 2010). Thirdly, a representation of the motor action cannot be evoked until the target object has been attended to. In action execution, the agent must activate a representation of the target object before its motor affordances can be computed (Johansson *et al.*, 2001); in action perception, the observer must compute the trajectory of the agent's hand onto the target (e.g. Oztop *et al.*, 2004). If these assumptions, which are individually quite well accepted, are brought together, an interesting model of event perception emerges, in which apprehending a transitive event involves a sequence of three SM operations: attention to the agent, then attention to the target, then activation of a motor programme. The idea that events have a characteristic temporal structure is certainly present in other models of event perception, in particular that of Reynolds *et al.* (2007). For Reynolds *et al.*, these sequen-

[1] Our terms 'agent' and 'patient' refer to Dowty's (1991) more general concepts 'proto-agent' and 'proto-patient'.

tial regularities relate primarily to the structure of an agent's *movements*: they are the kind of regularities that the 'biological motion' system becomes attuned to. In our model, such regularities are encoded within the action representation system, as discrete actions. But there is more to an event than an action. In our model, experiencing an event also involves a higher-level sequence, of relatively discrete *SM operations*. One of these is the activation of an action representation. But this operation must be *preceded* by an action of attention to the agent, and then an action of attention to the patient (if there is one). In our model, the notions of agent and patient are in fact *defined* by the serial order of attentional operrations in this SM sequence: the (proto-)agent is the first individual attended to; the (proto-)patient is the second.

Alongside this model of event perception, we also assume that the perception of each participant in an event involves its own canonically-structured sequence of SM operations. It is well established that in order to classify an object, an observer must first direct focal attention to the region of space it occupies. But observers can also attend to a region of space containing a homogeneous group of objects. Walles *et al.* (2014) argue that in between focal attention and object classification there is an intervening attentional operation that selects a spatial *scale* at which the classifier will be deployed, determining whether the classifier identifies the local or global form (Navon, 1977) of the attended stimulus. This operation determines whether a single individual is classified or a homogeneous group of individuals. In summary, perception of an individual involves a SM routine comprising three operations: selection of a salient region of space, then selection of a classification scale (determining whether a singular or plural stimulus will be classified), and finally activation of an object category. Event perception, in turn, is a higher-level sequential SM routine, some of whose elements have their own sequential structure.

WM representations as prepared sequences

We propose that representations in semantic WM exploit the sequential structure of perceptual processes. Specifically, we propose that WM representations of both individuals and episodes take the form of *prepared sequences*. This proposal is attractive for several reasons. For one thing, it offers a clear account of how semantic WM representations can influence SM processing: a prepared SM sequence is an 'executable' structure, that can initiate sequentially structured SM activity (including actions). For another thing, it suggests an account of a puzzling recent finding: stimuli held in WM appear to be transiently *reactivated* in SM areas during the delay period (see e.g. Meyers *et al.*, 2008). If WM representations are prepared SM routines, that can be executed in simulation, then active simulation processes could occur during the delay period, resulting in these transient patterns of SM activity. Finally, the proposal places semantic WM representations within a class of neural representation that is relatively well understood. We know a lot about how prepared sequences of attentional or motor operations are represented, particularly in macaques. The relevant representations are predominantly in prefrontal cortex (PFC), which is also a key site for semantic WM. A particularly interesting result is from Averbeck *et al.* (2002). They showed that the PFC assembly that stores a prepared sequence of SM operations contains sub-assemblies representing each individual operation—and moreover, that these sub-assemblies are active *in parallel* in the structure representing a planned sequence, even though they represent operations that are active one at a time. Our model will make use of this finding.

A model of semantic role-binding using sequentially-structured WM representations

Modelling semantic WM representations as prepared sequences suggests a novel account of how semantic roles are bound to participants in representations of events. Our account makes use of three ideas, which we introduce below.

The key idea is that the binding mechanism is implemented as part of the active process of rehearsing SM routines, rather than within a static representational structure. The classic binding problem arises because the SM media representing an individual's properties (location, shape etc) naturally represent just *one* individual: if the properties of several individuals are represented, it is hard to specify which properties belong to which individual. If a WM event representation supports the simulation of a sequential SM routine in which representations of agent and patient are active in these media *at different times*, many of these problems go away.

Of course, the event representation must still make reference to both participants, so it can activate these temporally separated representations. The second idea in our binding scheme is that event representations represent participants using *pointers* into the medium representing individuals—and that there are separate pointers for agent and patient. The pointers are active simultaneously in a WM event representation, but they are only *followed sequentially*, when an event is rehearsed. (Event representations thus conform to the properties of prepared sequences identified by Averbeck *et al.*) In neural networks terms, agent and patient are coded 'by place' in our WM event representations, in separate groups of units. Place coding of this kind is not normally seen as a viable way of implementing role-binding: a simple place-coding scheme suffers from the fact that there is nothing in common between representations of John-as-agent and John-as-patient. But if the place-coded representations of agent and patient just hold pointers into the medium representing individuals, which are activated at different times, this problem does not arise.

The third idea in our binding scheme is that the place-coded pointers in WM event representations do not point directly to SM media representing individuals, but rather to a *WM* medium holding representations of individuals. Recall that representations of individuals also have internal structure: we proposed above that the WM representation of an individual is also stored as a prepared, replayable SM routine.

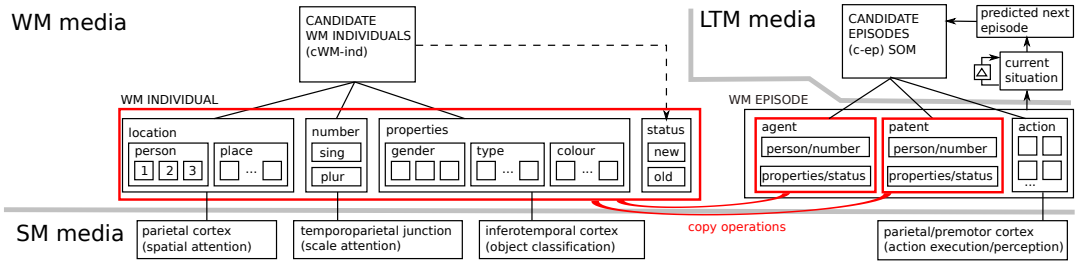


Figure 1: Architecture of the model of semantic WM

In our model, WM representations of recently-perceived individuals are held in a separate WM medium: the agent and patient representations in a WM event point to, and sequentially re-activate, representations within this WM medium. During rehearsal of a WM episode, these sequentially reactivated representations create opportunities for secondary rehearsal operations, simulating the steps involved in perceiving the participant individuals. This scheme introduces a measure of hierarchy in the model of role-binding, enabling the representations filling semantic roles to have a degree of internal structure—an important requirement, as noted earlier.

A model of WM episodes and WM individuals

Our model is illustrated in Figure 1. SM media are below the grey line; WM media are above it. The WM system representing individuals is on the left, and that representing events (or 'episodes', as we call them here) is on the right. The copy operations implementing pointers are highlighted in red.

The WM medium on the left holds a representation of a single selected individual, a **WM individual**, stored as a prepared sequence of a **location**, a **number** (i.e. classification scale) and a set of perceptual **properties**. These three representations are activated in parallel in the WM medium, but when the prepared sequence is executed or rehearsed, they activate associated first-order representations in the attentional and classification systems one at a time, as discussed above.

The media representing a WM individual provide input to another layer, the **candidate WM individuals** (**cWM-ind**) layer, which stores *combinations* of location, number and type over a short time period, and thus represents the set of recently-attended individuals. A partially specified WM individual can function as a query to the cWM-ind layer: if we specify a location, we may be able to retrieve an associated type and number (and vice versa). If we can, then the individual retrieved is classed as 'old'; if we cannot, it is classed as 'new'. These attributes are recorded in a **status** field of the WM individual, which is not part of the prepared sequence. Queries formed from partially-specified WM individuals can be used to generate expectations about the location or properties of individuals in the current scene, as we discuss below.

The **WM episodes** system is structurally similar to the WM individuals system. It holds a representation of a single selected episode, a 'WM episode', stored as a planned sequence of operations activating an **agent**, a **patient** and an **action**. As noted above, the agent and patient media hold content-addressed pointers to representations in the WM individual medium. All the media within a WM episode are active in parallel, but when a WM episode is executed or rehearsed, the representations they point to become active sequentially: the 'agent' and 'patient' media activate two successive representations in the WM individual medium, and then the 'action' medium activates a representation in the (pre)motor system.

The prepared operations in a WM episode also provide input to a layer holding episode representations learned over a longer timespan, the **candidate episodes** (**c-ep**) layer. This layer is a self-organising map (**SOM**): when exposed to training episodes, it learns to represent episodes as localist units, organised so that similar episodes are close together in the map. Each unit can encode a particular combination of representations in the agent, patient and action media, and thus can represent a complete episode by itself. Note this localist scheme is enabled by our model of binding: the 'agent' and 'patient' fields of a WM episode index their fillers *by place*, so carry information about both roles and fillers. Clearly, we cannot represent every possible episode using localist units. But that is not the purpose of the c-ep SOM: its role is rather to represent the episodes that occur frequently, so these can provide a top-down bias on SM processing during experience. Since the c-ep SOM uses localist representations, it can also represent *multiple* expected episodes simultaneously: a useful property, as we will show.

A final component of the network is a layer representing the **current situation**. In our model, this is the hidden layer of a recurrent network that learns to predict the *next episode*, given the episode that has just occurred, plus a copy of its hidden layer at the previous time point. The current situation network learns to predict a *distribution* of possible next episodes in the c-ep SOM (exploiting its ability to represent multiple episodes). (It is somewhat analogous to Reynolds *et al.*'s (2007) recurrent network for event representation, but Reyonlds *et al.*'s network predicts the next *component* of an episode, rather than the next discrete episode.)

One useful feature of our model is that that the c-ep SOM can learn *generalisations* over episodes. One kind of generalisation is hard coded in the model: the copies of WM individuals created in the agent and patient fields ignore location information, so representations of episodes in the SOM abstract over the location of participants. In our model, expectations

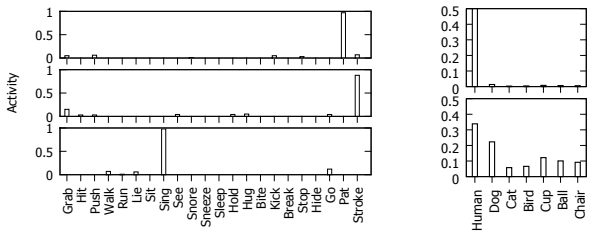
about the locations of objects are dealt with in the WM individuals system, as we will illustrate below. This step considerably reduces the combinatorial possibilities that need to be represented in the SOM. But the SOM also learns generalisations of its own. The ability to generalise is a standard feature of learning in SOMs, since episodes that are sufficiently similar will activate the same localist unit. In particular, since the representations of agents and patients providing input to the SOM are distributed, the SOM can learn to abstract away from the properties of token individuals and represent the participants of episodes as types, as we will show.

Here are some technical details about the network's architecture. The WM individual layer consists of localist sets of feature units for person (1/2/3), number (Sg/Pl), gender (Male/Female/Neuter) and status (new/old). Each set of units can either encode a single property unambiguously, or a probability distribution over properties. The type area also contains feature sets coding animacy, and object type (person/dog/cat/bird/cup/ball/chair). Object location, situated on a 100×100 grid, is coded by a population of 6×6 neurons with Gaussian receptive fields evenly covering the grid. Colour is coded by a population of 11 neurons with Gaussian receptive fields in 3D RGB space maximally responding to standard 11 basic colours (see Figure 3a). Such population coding is neurally plausible and there is a straightforward mathematical way of computing the likelihoods of different stimuli given the activities of neurons in the population (Jazayeri and Movshon, 2006). SM representations (below the grey line in Figure 1) are isomorphic to the WM areas they interface with. Likewise, agent and patient layers of a WM episode are isomorphic to the relevant parts of a WM individual. The 'action' area consists of 22 localist units for actions (see the *x*-axis legend in Figure 2a) and 11 units for their distributed featural codes. The cWM-ind layer is a variable-sized convergence zone of units fully connected with the WM individual layer. When a novel candidate individual is encountered, a new unit in the cWM-ind layer is recruited and the current values of WM individual units are copied into its connection weights (one-shot learning). The c-ep layer is a SOM with 400 units. Each unit also maintains a scalar weight reflecting the frequency of 'hits' for this unit, i.e. the number of times it was the most active unit. These frequency weights serve as priors for computing the Bayesian probability that the current input corresponds to an episode represented by a particular unit (for details see Takac and Knott, 2016). The network that represents the current situation is a recurrent SOM (a 'Merge SOM', Strickert and Hammer, 2005). This provides input to a layer of linear perceptrons which are trained to predict the next episode. Details of all these networks can be found in Takac and Knott (2016).

Training and testing of the network

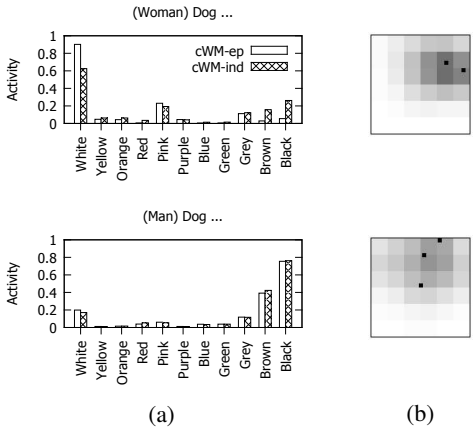
Training We trained the network by simulating SM experience of a sequence of episodes. Each episode is represented in the SM system as a complex sequence of SM operations. We begin by describing the properties of the individuals that featured in episodes. We created a fixed population of token individuals: each with a type, a number, and location and colour properties that are stochastically chosen based on its type. Locations are quasi-randomly generated as positions on a 100×100 grid (which in the system's 6×6 location medium are represented using coarse coding). Colours are stochastically generated from Gaussian distributions centred on 11 basic colours. We then generated a stream of episodes involving these token individuals. Each episode is presented to the WM system as a sequence of SM input items. Episodes are of three types: **transitive** (agent→patient→trans-action), **intransitive** (agent→intrans-action) and **causative** (agent→patient→cause-signal→unaccusative-action). In each case the agent and patient signals has a sequential structure of their own, namely location→number→type/properties. Each of these latter sequences is sent to the WM individuals medium, activating the different components of a WM individual representation one by one. When complete, the WM individual was first passed as a query to the cWM-ind layer, to find out whether the individual it represents has already been encountered. For each candidate unit currently active in the cWM-ind layer, we compute the likelihood that it corresponds to the current stimulus in the WM individual (Jazayeri and Movshon, 2006). This reduces to the average pairwise Kullback-Leibler divergence between the respective areas of the WM individual and the candidate unit weights (see Takac and Knott, 2016 for details). If a likely-enough candidate is returned, it is updated if necessary and the WM individual's status is set to 'old'; otherwise a new entry in the layer is created and the WM individual's status is set to 'new' (candidate units that have not been updated for N episodes are removed). The WM individual is then copied (along with its status) to the appropriate layer in the WM episode medium: either the 'agent' layer or the 'patient' layer. When a complete episode has been presented to the system, the layers in the WM episode are passed as input to the c-ep SOM. This SOM learns in the standard way. Note that while learning in the cWM-ind layer happens in a 'one-shot' manner, it only happens gradually in the c-ep SOM, current situation and next episode media.

Testing the sequence-based binding scheme To demonstrate the new binding scheme, we must show how the WM representations created during experience of an episode allow it to be *replayed*. To test this, after each episode is presented, the WM episode medium is used as input to a replay process, in which the layers in this medium activate the representations they point to one by one. When the activated representations are in the WM individual layer, they are used as a query to retrieve a location representation (recall that the location of individuals is not copied to the WM episode). If the binding scheme is effective, this process should regenerate the same sequence of first-order SM signals that was presented to the network during experience of the episode. In our tests, the sequence was perfectly reconstructed for 99.6% of episodes; this shows that our proposed binding mechanism is effective.



(a) (b)

Figure 2: (a) Action types predicted in the c-ep layer for 3 episode fragments. From top to bottom: man→dog→?, man→cat→?, bird→?. (b) Patient types predicted for Sequence *A* (top) and Sequence *B* (bottom).



(a) (b)

Figure 3: (a) Expectations about the colour of the patient generated by the c-ep and cWM-ind layers for episodes woman→dog (top) and man→dog (bottom). (b) Expectations about location of the patient generated by the cWM-ind layer for these episodes. Darker areas mean stronger expectations. Black dots represent actual locations of currently present white dogs (top) and black dogs (bottom).

Testing the network's prediction/generalisation abilities The network can make several kinds of prediction; we will focus on three progressively more complex predictions. Firstly, the c-ep SOM can make predictions about the episode currently being experienced, as experience is under way. Predictions about actions are easiest to demonstrate, since it represents actions directly. To evaluate these predictions, we introduced some regularities in the episodes presented to the system. Birds always sang (bird→sing); also when people interacted with dogs and cats, they always patted dogs and stroked cats (person→dog→pat, person→cat→stroke). We presented the c-ep SOM with episodes involving these participants, leaving the action field blank, and generated a distribution of expected episodes in the SOM; from this we reconstructed a distribution of expected actions, by linear combination of the weight vectors of SOM units, weighted by unit activity. Figure 2a shows these distributions are correctly weighted towards the actions encountered during training.

The c-ep SOM can also make predictions about the agents and patients of episodes. These are more complex, because its predictions are relayed to the WM individuals system, which refines them based on its own knowledge. To illustrate this process, we introduced a further regularity into the training episodes: in all episodes involving people interacting with dogs, the dogs were black if the agent was a man, and white if the agent was a woman; additionally, people always appeared in the top-left quadrant of the spatial array, and animals in the top-right quadrant. We then generated an underspecified representation in the WM episode: in the agent part, we activated a representation of a person (either man or woman), and in the patient part we activated the type 'dog', unspecified for colour. We used this representation to generate a distribution in the c-ep SOM, from which we reconstructed a predicted distribution of patient features. This whole distribution was copied to the WM individual medium, where it activated a distribution of units in the cWM-ind network. This distribution was used top-down to reconstruct a distribution over expected locations for the patient individual, and to refine the distribution over expected types and properties. Figure 3a shows activity in the colour-coding features of the resulting WM individual expectation. The system correctly predicts a colour centred on black in RGB space for man→dog episodes, and on white for woman→dog episodes. Importantly, the cWM-ind layer can also generate expectations about the location of the dog—see Figure 3b. There is a general bias towards the quadrant containing animals, since dogs always appear in this quadrant. But there are also specific biases towards the location of the black or white dogs that the system has recently encountered, that are *based* on its expectations about the colour of the patient dog.

Finally, the network can make predictions about the *next* episode, using its representation of the current situation. To show this, we presented it with a sequence of training episodes, with constraints on transitions between episodes: when a person hit a dog and then patted the same dog ('sequence *A*'), the dog always bit the person; when a person patted a dog after some other episode ('sequence *B*'), any random episode could follow. We tested the network by presenting sequences *A* and *B*, propagating activity through the Current situation and Next episode prediction layers to obtain a prior distribution over predicted next episodes in the c-ep layer. From this we reconstructed an expected distribution of agents, patients and actions in the WM episode. The network correctly predicted 'dog-bite-man' after Sequence *A*, and made a more neutral prediction after Sequence *B* (see Figure 2b). The learned update rule encodes something like 'If a man hits a dog then pats it, *the dog bites the man*.'

Roles for the network in language processing

As discussed at the outset, semantic WM representations must do service in language processing as well as in SM processing. Our WM model was designed with this in mind.

We envisage several linguistic roles for the network. The cWM-ind medium can function as the medium storing salient discourse referents, and the current situation medium can be understood as holding a representation of the current discourse context; for details on these ideas, see Takac and Knott (2016). We also envisage that the WM model plays a role in sentence processing. Specifically, we propose that generating a sentence reporting an episode stored in WM involves *rehearsing* this episode, in a special cognitive mode in which SM/WM representations can trigger output phonology. We built a neural network model of sentence generation implementing this idea (Takac *et al.*, 2012). Within this model, we argue that the syntactic concept of a **head** can actually be derived from the architecture of the semantic WM system. In syntactic theory, information conveyed by a syntactic head spreads through its local syntactic domain (e.g. a clause or noun phrase (NP)). This spreading process is seen most clearly in agreement phenomena: for instance, subject-verb agreement within a clause, or determiner-noun agreement within a NP). In a NP, agreement rules relate to the head features **person**, **number**, **type**, semantic **gender** and **definiteness**: exactly the information that is maintained tonically in a WM individual during its rehearsal.[2] Heads in a clause can convey all this information about the subject and object, and additionally information about the type of the action: exactly the type of information maintained tonically in a WM episode during its rehearsal. We argue that syntactic heads have an extended syntactic domain because they are read from WM representations that are tonically active during rehearsal, and so can influence phonology at multiple points. Again see Takac and Knott (2016) for further discussion.

Discussion

In this paper we propose that events and their participants are represented in WM as prepared SM sequences. This has several benefits. It permits a direct account of how WM event representations are created during experience, and of how they in turn influence event perception by generating top-down expectations. It enables a new model of role-binding, that allows hierarchical representations of event participants, and localist representations of candidate events, through a novel use of indexing and place-coding. This in turn allows the model to represent large *distributions* of expected events: a very useful ability, which the models of van der Velde and de Kamps (2006) and Stewart and Eliasmith (2012) do not have. Finally, the network supports several aspects of a model of language processing; most interestingly, it creates a framework within which aspects of syntactic structure can seen as deriving from the structure of the semantic WM system.

The model as illustrated here leaves many open questions. One question concerns space requirements. Our place-coding scheme for event participants requires creating several separate *copies* of the WM individual medium, which is expensive as regards storage space. However, storage is within acceptable limits when scaled up to a memory of realistic size, as discussed in Takac and Knott (2016). Another question concerns recursive representations. While the model described here implements a notion of *hierarchical* representations, these representations are not recursive. Again, this issue is addressed in Takac and Knott (2016), where we argue that the current scheme extends well to an account of complement clauses, subordinate clauses, and relative clauses.

[2] 'Person' is conveyed as a special kind of location information: the location can be a point in external space (in which case the point is also specified) but it can also be a direct reference to the speaker or interlocutor, whose location is presumed to be independently given. Type and semantic gender are expressed within the complex of perceptual properties. Definiteness is expressed by status (new or old).

Acknowledgements

This work was supported by the New Zealand Marsden Foundation through Grant 13-UOO-048. Martin Takac was also partially supported by Slovak KEGA grant 017UK-4/2016.

References

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

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

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

Jazayeri, M. and Movshon, A. (2006). Optimal representation of sensory information by neural populations. *Nature Neuroscience*, **9**(5), 690–696.

Johansson, R., Westling, G., Backstrom, A., and Flanagan, J. (2001). Eye-hand coordination in object manipulation. *Journal of Neuroscience*, **21**(17), 6917–6932.

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

Meyers, E., Freedman, D., and Kreiman, G. *et al*. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J.Neurophys*, **100**, 1407–19.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, **9**, 353–383.

Oztop, E., Bradley, N., and Arbib, M. (2004). Infant grasp learning: a computational model. *Exp. Brain Research*, **158**, 480–503.

Reynolds, J., Zacks, J., and Braver, T. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, **31**, 613–643.

Stewart, T. and Eliasmith, C. (2012). Compositionality and biologically plausible models. In *The Oxford Handbook of Compositionality*. Oxford University Press, New York.

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

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.

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 der Velde, F. and de Kamps, M. (2006). Neural blackboard architectures of combinatorial structures in cognition. *Behavioral and Brain Sciences*, **29**, 37–108.

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

Webb, A., Knott, A., and MacAskill, M. (2010). Eye movements during transitive action observation have sequential structure. *Acta Psychologica*, **133**, 51–56.