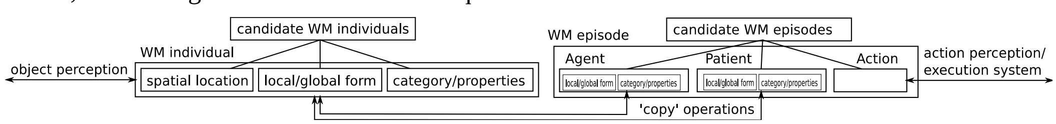
A model of semantic working memory: representing episodes and individuals as prepared sensorimotor sequences

Alistair Knott[1] (alik@cs.otago.ac.nz), Lubica Benuskova[1], Martin Takac[1,2],
[1] Dept of Computer Science, University of Otago; [2] Centre for Cognitive Science, Comenius University, Slovakia

In this paper we present a neural network model of semantic representations in working memory (WM). There is good evidence that the episodes we encounter during sensorimotor (SM) interaction with the world are stored in WM as well as in long-term memory (LTM). Experiencing an episode takes at least a second, often longer, but to store an episode in the hippocampus requires its component elements to be active within around 100ms (the time window for LTP), so episodes must be buffered in WM before being encoded in the hippocampus. There is also evidence for a semantic WM medium in language processing studies. For instance, WM for word sequences is much better if the words form a sentence, even for patients with hippocampal damage: the standard explanation is that the meaning of a sentence is stored in WM separately from its phonology, and provides an additional cue to recall.

Our network model of semantic WM highlights its twin roles in linking the SM and LTM systems and in interfacing to language. Our starting assumption is that that the sequential structure of SM processes provides an organizing principle for semantic WM representations. We have studied the SM routines through which simple transitive actions are experienced, and found evidence these routines have a well-defined canonical sequential structure, whereby the agent is identified first, and then the patient, and then the action. We have also found that the process of identifying individual participants in episodes has its own well-defined sequential structure, involving first attention to a location in space, then attention to either the local or global form of the stimulus at this location, and finally classification of this form. In our model, both episodes and the individuals that participate in them are represented in WM as *prepared SM routines*, in the same areas of prefrontal cortex (PFC) that hold prepared sequences of motor/attentional actions. The way prepared sequences of actions are stored in PFC is relatively well understood. Of particular interest is the finding that the component actions of a prepared sequence are active *in parallel* in PFC, even though the actions occur in sequence. The architecture of our model is shown below.



The **WM individual** (bottom left) holds a representation of the currently attended individual, and the **WM episode** (bottom right) holds a representation of the episode currently being experienced. Although perceiving both individuals and episodes involves a sequence of SM operations, both WM media represent their component operations in parallel. A key idea in our model is that perceiving an episode involves perceiving its participant individuals sequentially. The WM individual medium first represents the agent of the episode, and then the patient. However, these representations are *copied* to different locations in the WM episode medium, which maintains these copies in parallel. While the WM episode holds 'place-coded' representations of agent and patient, its participants are represented the same way regardless of the role they play, because they are copied from the same individual-denoting medium. (NB the copy operations are only possible if an episode's participants are perceived sequentially.)

Our model of semantic WM representations is novel in three respects. First, since both individuals and episodes are represented as parallel patterns of activity, the model can learn *probability distributions* over possible individuals and episodes. The **candidate WM individuals** medium learns to associate spatial locations with the properties of individuals recently encountered at those locations: when cued with a location, it evokes a distribution over likely object properties, and vice versa. The **candidate WM episodes** medium learns localist representations of whole episodes, using a self-organising map architecture: when cued with a partial episode representation, it evokes a distribution of full episodes consistent with the cue. During episode perception, this distribution is updated online as new components of the episode are perceived. Second, the model gives a novel account of how episodes in WM are stored in the hippocampus. In our model, WM episode representations are 'executables', which can be internally *rehearsed*, at speeds commensurate with hippocampal LTP: when thus rehearsed they can be encoded as sequences in the hippocampus. Finally, our model does service in an account of language processing. WM episodes can also be rehearsed in a special 'language mode', in which SM signals activate output phonology: here, a rehearsed sequence of SM signals leads to the production of a sequence of words. In this mode, the model can learn the syntax of many languages, including English, Slovak and Māori. The fact that semantic WM representations encoding SM sequences can support a model of syntactic development raises the possibility that the 'universal syntactic structures' posited by Chomsky may have their origins in the structure of the SM and semantic WM systems.