A computational exploration on the role of semantic memory in episodic future thinking

Yuichi Ito (ito.yuichi@nagoya-u.jp), Taiji Ueno (taijiueno7@gmail.com),
Shinji Kitagami (kitagami@cc.nagoya-u.ac.jp), and Jun Kawaguchi (kawaguchijun@nagoya-u.jp)
Department of Psychology, Graduate School of Environmental Studies, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya City, Aichi 4648601, JAPAN

Abstract

Episodic future thinking refers to a human cognitive process which generates successive predictions of events that are likely to occur in a cue-specific context in the future. An emerging view is that semantic memory as well as episodic memory contributes to this process, but the exact mechanism remains unclear. We built a computational model that learned to predict the next event upon a presented event (sequence prediction model). After learning the statistical structure in the training sequence, the model was tested for generating successive self-predictions of events triggered by a cue. The generated sequence of events captured some phenomenological features of patients with semantic dementia when the semantic system of the model was damaged. The role of semantics in episodic future thinking and the usefulness of a sequence prediction model are discussed.

Keywords: episodic future thinking; semantics; parallel-distributed processing model; sequence learning

Introduction

We can project ourselves into the future despite the fact that we have never experienced it. The term *episodic future thinking* refers to a human ability to envision a plausible future event in a specific time and place (i.e., a specific context) (Atance & O'Neill, 2001; Schacter, Addis & Buckner, 2008). Over the last decade, researchers from various fields, including psychology, neuropsychology, and neuroimaging, have investigated episodic future thinking, focusing mainly on the contribution of episodic memory to constructing episodic future thought. More recently, data from patients with semantic dementia have suggested that semantic memory may also play a role (Irish, Addis, Hodges, & Piguet, 2012). The current study used a computational model to investigate the mechanism by which semantic memory supports episodic future thinking.

Role of Episodic Memory

The role of episodic memory has been suggested in various studies. For example, some neuroimaging studies have revealed a common neural network involved in the remembering of past, and in imagining future events (Szpunar, Watson, & McDermott, 2007). These data are consistent with neuropsychological studies with amnesic patients (e.g., hippocampal amnesia or Alzheimer's disease) who showed simultaneous impairments in both remembering past episodes and imagining future events (e.g., Irish et al., 2012). Based on these findings, the *constructive episodic simulation hypothesis* was proposed, which assumes that imagining future events requires a system that can retrieve detailed information stored in episodic memory and flexibly recombine them into *coherent* representations of future events (Schacter, Addis, & Buckner, 2008). Further support for this idea comes from experimental psychology. For example, both retrieving an episode and imagining a future event are affected by a temporal distance factor in the same manner. Specifically, Addis, Wong, & Schacter (2008) collected both the past events that participants recalled and the future events they generated, and classified detailed information in these outputs as either *internal* or *external*. Internal details are "episodic" information, meaning specific in time and place and related to the central events (i.e., the main event described by the participant). In contrast, external details are not specific in time and place. It was found that, in both recalling of past episodes and thoughts about future episodes, internal details lessened as participants were required to produce farther events from the present in both directions. This means that as episodic future thinking goes farther in terms of temporal distance from the present, the time and place (context) of the generated events deviates from those of the central events (central topic).

Role of Semantic Memory

More recently, the role of semantic memory in episodic future thinking has also captured attention (Irish et al., 2012). D'Argembeau and Mathy (2011) suggest that construction of episodic future thought typically involves progressive conversion from general to more specific information such that access to general knowledge (semantics) precedes retrieval of time-specific episodic information. In other words, semantic memory provides a "framework" for construction of episodic future event representations, and then episodic information from the past is integrated to form a coherent and elaborated sequence of future events. A key support for this idea came from a study with neurological patients with semantic dementia, characterized by the progressive and insidious loss of conceptual knowledge about objects, facts and the meaning of words, yet preserved non-verbal episodic memory (Irish et al., 2012). Specifically, Irish et al. (2012) found that although their patients were as good at remembering past episodes as controls, their episodic future thoughts lacked internal details. In other words, the sequence of events they generated did not maintain the time and place information (context) that was cued by an investigator. Note that this was not due to a difference in task difficulty because Alzheimer's disease patients in this study showed simultaneous impairments in both measures. Thus, this dissociative pattern suggests that even if episodic memory is relatively intact, loss of conceptual knowledge has an impact on episodic future thinking.

Motivated by these findings, we employed a parallel-distributed processing (PDP) modelling approach to investigate the mechanism by which semantic memory contributed to episodic future thinking. As we reviewed above, human experiments have provided significant insights, but each has its own limitation: It is relatively difficult to separate the contribution of episodic memory from that of semantic memory in healthy controls. Semantic dementia patients are the best test cases but their verbal outputs are limited such that it is difficult to probe their cognitive processing in detail. In contrast, computational modelling provides an ideal situation where we can directly look at the nature of computation/representations in the model to glean further insights into how semantic memory supports other cognitive processing (e.g., Woollams, Joanisse, & Patterson, 2009).

Future Prediction Model

Given there is no computational model for episodic future thinking in the literature, the initial step is to make some simplified assumptions so that the target cognition can be implemented in a computational model. A standard paradigm to probe episodic future thinking involves a presentation of a cue such as time/location/object (e.g., next year's birthday, or 50th birthday, etc.), and a participant successively generates cue-specific predictions on what is likely to happen (e.g., a birthday cake is on a plate in a dining room → I blow the candle → my friend will pick out the candle → the friend will cut the cake → the friend will serve me a cake on a plate, etc.). The nature of this generation is *successive* such that the order of these example sentences cannot be at random. In other words, future thinking includes at least two aspects - computing cue-specific information and successively generating future predictions based on the corresponding previous prediction. Of course, these two aspects are not enough to account for the whole episodic future thinking processing. However, once we assume that episodic future thinking taps at least an ability to generate successive predictions based on the corresponding previous prediction upon a time-/location-/object-specific cue, then there is an existing computational model by Elman (1990) that we can adopt and modify for the current purpose. This model was trained for predicting the next alphabetic letter in an artificial language. Specifically, the model received a 6-bit binary vector, which represented one of the alphabetic letters, and the model was trained for predicting the next 6-bit binary input vector. The presented sequence was not random, but there was a statistical structure regarding what was likely to come next (artificial *grammar*). The model learned this statistical structure in the sequence. In later studies, human participants were trained for the same task and were able to use their statistical knowledge after training in order to generate successive predictions about the next letter following their own previous predictions upon a presented cue (Perruchet & Amorim, 1992). Returning back to the current study, it would be possible to assume that a statistical structure exists even in the event sequence (episode) within the real world. For example, we reasonably guess that the next event would be to blow the candle when a birthday cake is served to the dinner table. Also, we know that someone will cut cake into pieces before biting into a whole cake. Thus, there is some statistical structure in the *sequence of events in real world*, and our working hypothesis is that the order of successive cue-specific predictions in episodic future thinking should be to some extent constrained by this statistical structure in real world. Once we assume the similarity between the future prediction of the next letter in a given language (Elman, 1990) and the future prediction of the next event in real life, then it is natural to adopt Elman's approach for modelling episodic future thinking (see below in detail). As we admit above, episodic future thinking is a complex cognitive process, but this approach is promising to capture at least the two core characteristics of episodic future thinking mentioned above.

Method

Model Architecture, Tasks, and Representations

Figure 1 shows the architecture of the model. Four peripheral layers (input layer, output layer, semantic layer, and recognition layer) were connected bidirectionally through a single hidden layer. The hidden layer and each of the five output layers were connected to themselves. The input layer was sub-divided into five layers, each of which represented one of the five elements of the current event (Figure 1). For example, the first layer represented the context information of the current event. If this context layer was hard-clamped to the binary vector of [1 0 0 0 0 0], then it meant the current event occurred in Context 1 (e.g., *school*). The remaining four layers represented the Agent/Action/Object/Instrument of the current event. Thus, if the whole input layer was hard-clamped to the 18-bit

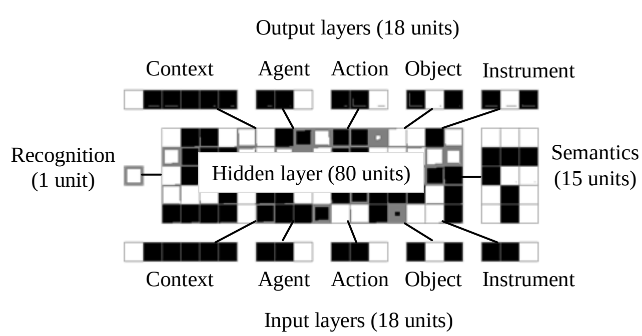

Figure 1: The architecture of the model (Hinton diagram).

Table 1: Sequence structure of the training set.

sequence	context			other information					
	context label	pattern	predictability	pattern				predictability	
				agent	action	object	instrument	with context	without context
event (1)	1	1 0 0 0 0 0	constant	0 1 0	1 0 0	0 1 0	1 0 0	17, 83, or 100%	6 ~ 45%
event (2)				1 0 0	0 1 0	0 0 1	1 0 0		
⋮					⋮				
event (i)				0 1 0	0 0 1	0 1 0	0 1 0		
event (i + 1)	6	0 0 0 0 0 1	constant	0 0 1	0 1 0	0 0 1	0 1 0	33, 50, or 100%	
⋮					⋮				
event (j)				1 0 0	1 0 0	1 0 0	0 1 0		
event (j + 1)	3	0 0 1 0 0 0	constant	0 0 1	0 0 1	0 0 1	0 0 1	100%	
event (j + 2)				0 1 0	0 1 0	1 0 0	1 0 0		
⋮					⋮				

vector of [(Context) 0 1 0 0 0 0 (Agent) 1 0 0 (Action) 0 0 1 (Object) 1 0 0 (Instrument) 0 0 1], then the current event was '*In Context 2 (e.g., home), Agent 1 (e.g., John) did Action 3 (e.g., cut) to Object 1 (e.g., cake) with Instrument 3 (e.g., knife)'*. The layers in the output side had the same structure, and when presented with the input pattern of the current event, the model was trained to activate the units in the output layer that consisted of the next event (the input 18-bit vector of the next trial). The sequence structure will be explained later.

Next, the semantic layer consisted of 15 units whose activation patterns represented the 'conceptual knowledge' of the current event (interpretation of the event) in a distributed manner. Following many parallel-distributed processing (PDP) models that incorporated a 'conceptual knowledge' system in their models (Woollams et al., 2009), no attempt was made to design semantic representations that captured the actual meanings of the input pattern (e.g., input words, action, event, etc.). Instead, like past models, artificial semantic representations were created that, nonetheless, captured core characteristics of the meaning of an event. Specifically, we assumed that the meaning of an event would be to some extent related to the action, instrument, and object information of that event (e.g., not an arbitrary mapping). Once we hear these pieces of information, we can guess what happened in that event with some confidence. In contrast, the meaning of an event would be less strongly related to information on who (Agent) did that action. For example, the meaning of *cutting an apple with a knife* is invariant irrespective of who did that action. Next, the context information also constrains the meaning of the event. We know that certain kinds of events rarely occur in a certain context. For example, *passing a ball* should not occur in a restaurant. Of course, Agent information would also constrain the meaning of an event (e.g., we might know that John would never eat an apple), but to a lesser extent than context/action/object/instrument information. Taking these assumptions together, we created the target semantic representations such that the bit-patterns in the context/action/object/instrument input layers were systematically related to part of the target vectors in the semantic layer (i.e., mapping was not completely arbitrary). Then, when presented with the current event pattern in the input layer, the network was trained for generating the correct pattern in the semantic layer in addition to predicting the next event in the output layer. Irish et al. (2012) demonstrated that semantic dementia patients were less accurate than controls for 'knowing (semantic)' non-personal events over the past/future 10 years. Thus, we damaged this layer in simulation of the patients' behaviour.

A recognition trial was occasionally inserted during training, in which the network was trained for judging whether the presented event pattern had been experienced before or not. The single unit in the recognition layer served to represent the network's recognition judgment. Specifically, the input layer was hard-clamped to the value of an event representation, and then the network was trained to activate this recognition unit (1.0) if the presented event representation had appeared ('old') before, as part of the main task. In contrast, the recognition unit should be turned off (0.0) if the presented episode representation had never appeared before ('new').

Sequence Structure of the Training Set

Sequence Structure of Context Information The sequence in the main trial was semi-random. Table 1 shows the structure of the sequence. First, as the left half of Table 1 shows, the context information (i.e., first 6-bit of the 18-bit input vector) was kept constant for several successive events in order to mimic the real world, where we experience successive events in the same context then move to another one. By presenting the first 6-bit information in this way, we can more safely argue that this 6-bit information represents the context information of an event. Thus, the predictability of the next context information was 100% in most trials unless it was the boundary of a context-block. After several events, the context information changed into another context semi-randomly (33%-50% predictability).

Sequence Structure of Agent/Action/Object/Instrument Information The sequence of the remaining 12-bit information of an event was also semi random. There were 81 possible input patterns, formed by crossing 3 (Agent) by 3 (Action) by 3 (Object) by 3 (Instrument). When the context information was not considered, the predictability of the next event (i.e., next agent/action/object/instrument information) varied from 4% to 45% depending on a trial. When the context was considered together, the predictability increased such that it varied from 33% to 100% depending on a trial. We implemented the constraint from context information to mimic the real world. For example, it is more difficult to predict what will happen if we see a *ball bouncing* at a restaurant, but it is less difficult to predict at a park.

Recognition Trials After every nine trials for event prediction (and simultaneous computation of meaning), six trials were inserted to train the model for event recognition. The network received a 18-bit input pattern, and was required to judge whether or not this pattern had been presented before as part of the main task by activating/deactivating the recognition unit. In order not to bias the network's response, 'old' and 'new' trials were evenly distributed (3 trials, each) within each recognition block. The 'old' events were randomly sampled from the main training trials that the network had experienced during event prediction. The 'new' event-set was created in the following steps. First, when we had created the sequence of the main trials, we had ensured that not all the 81 possible input patterns (formed by combining agent, object, action, & instrument) appeared in every one of the 6 possible contexts. Specifically, in each context, 20-27 possible combination of agent/object/action/instrument information had been randomly sampled and removed from the training set such that these patterns never appeared in that particular context during the main task. These pre-removed patterns served as 'new' events. To be clear, it was possible that these patterns appeared in another context. For example, the network might have received the 18-bit vector of [1 0 0 0 0 0, 1 0 0, 1 0 0, 1 0 0, 1 0 0 (comma denotes the boundary of layers)] but not received that of [0 1 0 0 0 0, 1 0 0, 1 0 0, 1 0 0, 1 0 0]. Then, the network would have to activate the recognition unit when presented with the former pattern but would have to deactivate the same unit in the case of the latter. Thus, the network was trained for recognition of a particular event involving a particular context/agent/object/action/instrument. We also ensured that not all the possible 'old' trials and 'new' trials were presented during training, such that we were able to probe the generalization performance of the network to the untrained 'old'/'new' patterns.

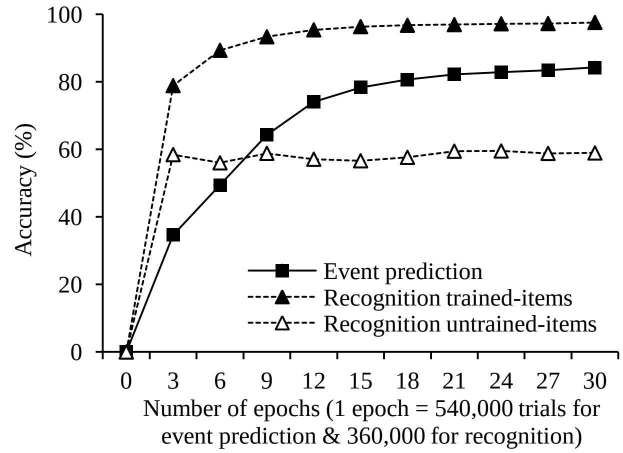

Figure 2: Learning curves for event prediction, recognition of trained-items, and recognition of untrained-items.

Training Parameters

In each trial, 18 units in the input layer were hard-clamped to their input values, and the network was allowed to cycle 10 times. In each time step, the activation spread to the next layer gradually being scaled by the values of the interconnecting weights, and the network settled into the steady state (called as an attractor). After 10 cycles of updates, the discrepancy between the output activation patterns (output event layer and semantic layer) generated by the network and the correct target pattern was calculated, and the connection strength was adjusted to reduce the discrepancy. In recognition trials, only the discrepancy in the recognition unit was considered. A learning rate of 0.01 was set at the beginning of the training. Then, every 10 epochs of training, the learning rate was gradually reduced by 0.001. A decay parameter was set to 0.0000001 at the beginning and gradually reduced by 0.00000001 as the learning rate was reduced. When we evaluated the network's performances during/after training, we used a strict criterion such that the output was scored correct if the discrepancy was within 0.5 in every unit of the target layer after the 10^{th} cycle (i.e., the activation is less/more than 0.5 if the target is 0.0/1,0. for each unit respectively).

Results

Trained Tasks

Figure 2 shows the learning curves for the event prediction task and the recognition task averaged across 10 independent simulations (initiated with different random seeds). The network successfully learned to predict the next event, thus acquiring the statistical structure which existed in the event sequence as well as recognizing the presented event pattern, which was generalized to untrained items. Accuracy for computing the meaning of an event quickly reached 100% after the training was initiated.

Episodic Future Thinking

As explained in the introduction, the current model focused to capture at least the ability to compute cue-specific events successively following its own previous event prediction, a core characteristic of episodic future thinking. Thus, we first presented cues (e.g., Context = *home*, Agent = *john*, Action = *cut*, Object = *cake*, Instrument = *knife*). Then, once the network generated an output (i.e., prediction of next event), we presented this output vector pattern as the input of the next event, and the network generated the next output (prediction of the next event following its own prediction, see Botvinick & Plaut, 2004, for the same approach in action learning). This cycle was reiterated 1000 times, and the generated 1000-event sequence was regarded as an approximation of the network's episodic future thinking. As a result, the network successfully kept the presented context information (Context 1) constant for the first 829 events (average of 10 simulations), but lost this context information after this point.

Simulation of Semantic Dementia

Following past simulations on semantic cognition, we simulated the episodic future thinking of patients with semantic dementia by removing some of the links between the semantic layer and the hidden layer (e.g., Woollams et al., 2009). Figure 3 shows how long (how many successive events) the network maintained the cued-context information as a function of disease severity (in terms of the proportion of links removed). This 'lesioning' simulation was reiterated 50 times with different links being sampled and removed, and the outcomes were averaged in order to avoid an idiosyncratic result. We found that, as the damage became more severe, the network was increasingly unable to maintain the event sequence of the cued-context (NB., The intact model kept the context for 829 events). Thus, future thinking deviated into another context/topic. Moreover, the proportion of the links removed was negatively correlated with the number of event predictions that maintained the cued-context [$r(17) = .-75$, $p < .01$], suggesting that semantics had a causal role in generating a coherent episode in future thinking. Importantly, event recognition accuracy was intact (more than 95% accurate) after this lesioning. All of these are consistent with the data from semantic dementia patients (Irish et al., 2012).

Discussion

The current model successfully acquired the statistical structure within the training set, and used this knowledge to generate a context-coherent sequence of events triggered by cues (episodic future thinking). Moreover, when the computation of semantic knowledge was impaired, the model could not generate a context-coherent event sequence, yet preserved its recognition ability of event patterns. Importantly, the number of the events generated in a specific context was negatively correlated with the severity of damage, suggesting the causal role of semantics in episodic

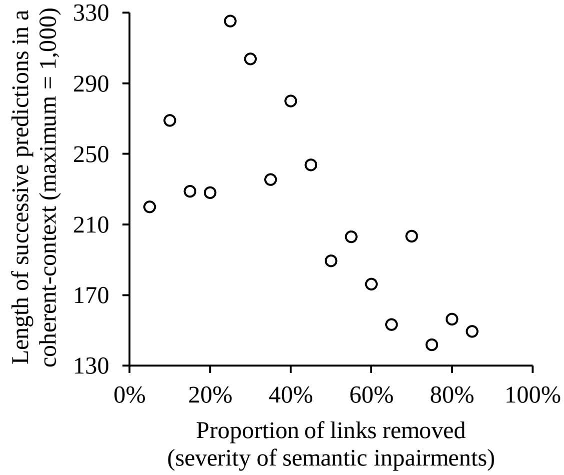

Figure 3: Numbers of successive events in which the network maintained the cued-context information as a function of disease severity.

future thinking (Irish et al., 2012). This is consistent with the idea that the semantic system provides the framework of the event (D'Argembeau & Mathy, 2011).

How does the semantic system affect the maintenance of context-coherent event sequences? This can be explained in terms of one of the general principles of PDP models. During training, a PDP network finds a unique attractor state (= unique abstract pattern in the hidden layer) associated with each of the input patterns. Once an input value is fed into the model, the activation gradually spreads, and the internal activity of the hidden layer gradually settles onto this unique status, as if it is *falling into its unique attractor basin*. They are unique, but similar inputs are associated with similar attractor basins. In the current model, the input patterns that share the same context information will fall into similar/neighbouring attractors, thus producing the same context output information to keep a context-coherent episode. However, if the internal representation of the model changes due to an impaired computation at some part of the model, then the network may settle into a wrong attractor basin, generating a wrong output. The diagnostic analysis suggests that this is certainly the case in our model. Specifically, we presented six events in different contexts to the network, and the activation pattern in the hidden layer on which the network settled was measured with/without semantics. Figure 4 shows the similarity structure of these patterns found by a multi-dimensional scaling analysis. With the intact semantic information (filled-markers), the network settles onto the context-specific attractor basins such that the network does not confuse one context with another. However, when the semantic system was damaged (open-markers), the network's internal status drifted away from its correct attractor, thus generating a different/wrong context representation (e.g., The open-circle is closer to the filled-diamond rather than filled-circle). In other words, semantic representations contribute to "binding" a time-varying event

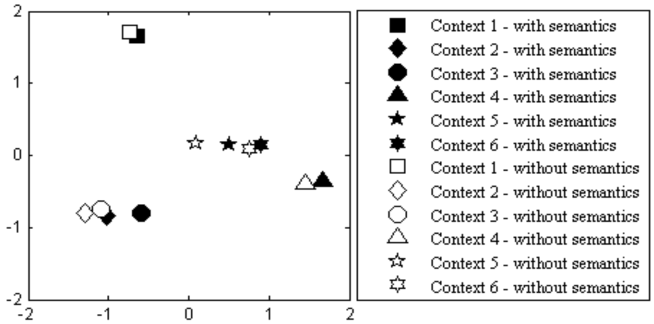

Figure 4: The similarity structure in the activation patterns of the hidden layer as a function of the input context information and of with/without semantics.

sequence such that it forms a context-coherent episode. One might describe this as a framework within which episodic details are integrated (D'Argembeau & Mathy, 2011). Interestingly, Schapiro et al. (2013) has recently demonstrated that temporally-close stimuli that form one coherent event are similarly represented (in terms of voxel-based neural patterns) in the inferior/superior anterior temporal lobe and inferior frontal gyrus, both of which are the damaged areas in semantic dementia patients. Damage in this area might disrupt in computation of such similar neural patterns, and bound stimuli might fall apart.

Then, the question is why collapsed semantic knowledge has little effect on episodic recognition accuracy, as was demonstrate in this model as well as in patients with semantic dementia (Irish et al., 2008). This is because recognition of a particular event is both context-specific and agent/action/object/instrument-specific. In other words, it is crucial not to confuse a new event with an old one, even if part of the information contained in that new event is semantically familiar (e.g., you have ever used that instrument before and/or have seen the same action conducted by the same agent, yet in a different context). Therefore, it is possible that event recognition is not influenced by degradation of semantic knowledge (or at least not detected with a standard test).

Admittedly, the ability to generate context-specific event predictions could be simulated if the modules representing schemas or scripts were explicitly built-in by a modeller a priori. However, the model implemented symbolic system must have assumptions about schematic knowledge preliminarily (further discussions, Botvinick & Plaut, 2004). The present sequential model did not have that symbolic system and developed by learning the statistical structure in the event sequence. This implies that learning sequential structure enables the model to compute schema-like representation (Botvinick & Plaut, 2004), and can capture the behaviour of semantic dementia patients..

In summary, we have clarified the mechanism by which semantics contribute to episodic future thinking. The sequence prediction model (Elman, 1990) is a useful computational framework that can be extended to an event sequence triggered by a cue such that it successfully captures the phenomenological and neuropsychological features of episodic future thinking. Certainly, this model does not capture the whole aspects of episodic future thinking, and in this sense, this is a proto-episodic future thinking model. In future work, implementation of essential factors for episodic future thinking is required such as the concepts of "self" or "temporal distance".

Acknowledgments

This study was supported by Grant-in-Aid for Challenging Exploratory Research (23653224) to J. Kawaguchi, and for JSPS Fellows (24008998) to Y. Ito.

Reference

Addis, D. R., Wong, A. T., & Schacter, D. L. (2008). Age-related changes in the episodic simulation of future events. *Psychological science*, 19, 33-41.

Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. *Trends in cognitive sciences*, 5**,** 533-539.

Botvinick, M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, 111, 395-429.

D'Argembeau, A., & Mathy, A. (2011). Tracking the construction of episodic future thoughts. *Journal of experimental psychology: General*, 140, 258-271.

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, 14, 179-211.

Irish, M., Addis, D. R., Hodges, J. R., & Piguet, O. (2012). Considering the role of semantic memory in episodic future thinking: evidence from semantic dementia. *Brain*, 135, 2178-2191.

Perruchet, P. & Amorim, M. A. (1992). Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 785-800.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature reviews Neuroscience*, 8, 657-661.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2008). Episodic simulation of future events: concepts, data, and applications. *Annals of the New York Academy of Sciences*, 1124, 39-60.

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013) Neural representations of events arise from temporal community structure. *Nature Neuroscience*, 16, 486-492.

Szpunar, K. K., Watson, J. M., & McDermott, K. B. (2007). Neural substrates of envisioning the future. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 642-647.

Woollams, A. M., Joanisse, M., & Patterson, K. (2009). Past-tense generation from form versus meaning: Behavioural data and simulation evidence. *Journal of Memory and Language*, 61, 55-76.