Emergence of Semantic Memory through Sequential Event Prediction and Its Role in Episodic Future Thinking: A Computational Exploration

Yuichi Ito (ito.yuichi@nagoya-u.jp) **Taiji Ueno (taijiueno7@gmail.com)**

Shinji Kitagami (kitagami@cc.nagoya-u.ac.jp) **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

This study aimed to clarify the mechanism underlying episodic future thinking, which refers to the ability to generate prospective events in a specific time/location/context. Given that episodic future thinking involves generating predictions in a plausible order from previous internal predictions, we hypothesized that knowledge of sequential event prediction should underlie episodic future thinking. A parallel-distributed processing model was trained to predict the next event in the training sequence. After training, the model used the acquired knowledge to repeatedly self-generate event sequences (i.e., the model predicts the next event, and this prediction then forms the input of the next trial which in turn will trigger the next prediction). The resultant event sequences captured the episodic future thinking of normal participants and that of neurological patients when the model was lesioned. Moreover, the nature of knowledge acquired after training for sequential prediction of external events reflected that of episodic memory, schema-like knowledge and semantic memory, all of which have been found to contribute to episodic future thinking by past studies.

Keywords: episodic future thinking; semantic memory; parallel distributed processing model; sequential prediction

Introduction

We can mentally simulate future events that are likely to happen in a specific time and place (e.g., "We'll go to that Indian restaurant for lunch. Upon arrival, a young waiter will say hello and show us to our table"). This cognitive function is termed *episodic future thinking* (e.g., Schacter, Addis & Buckner, 2008), and past studies have investigated the role of various types of knowledge within this mental simulation process. These include episodic memory (Hassabis et al., 2007), autobiographical memory (D'Argembeau & Mathy, 2011), schema/schemata representations (general knowledge database about a location/context where a mental simulation is projected) (Berntsen & Bohn, 2010), semantic memory[1] (Irish et al., 2012), and so on. For example, D'Argembeau and Mathy (2011) argued that construction of future event representations typically involves gradual conversion from general to more specific information such that access to general knowledge (autobiographic memory, and schema) precedes retrieval of time-specific episodic information. The role of semantic memory is also supported by neuropsychological data from patients with semantic dementia (SD) (Irish et al., 2012). These patients are impaired on tasks that probe conceptual knowledge of things (words, object, etc.), but their episodic memory (especially, about recent events) are relatively preserved (Irish et al., 2011). Irish et al. (2012) revealed that their episodic future thoughts lacked details relevant to the events cued by investigators. For example, when asked to talk about a future dinner, an SD patient might suddenly change topic and talk about his wife or past events. In other words, SD patients' future simulations tend to *transgress the boundary of the contexts* cued by investigators. However, it has yet to be clarified why and how these different types of representations contribute to the simulation of future events. An implemented computational model is a useful approach on this issue (e.g., Botvinick & Plaut, 2004; Elman, 1990).

Any computational modelling of a complex higher-order cognitive function requires a set of working assumptions and simplifications. It is noteworthy that simulation of future events involves the *self-generation of successive internal event predictions*. Taking a simulation of tomorrow morning as an example, one might first envision waking up in your bedroom, followed by an image of a next plausible event such as leaving the room, and finally one might imagine washing his/her face. In other words, our working assumption is that mental simulation requires a mechanism that allows sequential prediction of an event after the previously self-generated event in a plausible order. Interestingly, a seminal work of Rumelhart et al. (1986) mentioned this idea more than a quarter-century ago:

> *Now, suppose that the world events did not happen. It would be possible to take the output of the mental model and replace the stimulus inputs from the world with inputs from our model of the world. In this case, we could expect that we could "run a mental simulation" and imagine the events that would take place in the world when we performed a particular action.*

Sequential predictions and various knowledge

Once we formulate a mental simulation of future events in terms of sequential predictions (NB. We do not mean episodic future thinking is *equal* to sequential predictions), then we can explain why the experimental studies above found correlations between episodic future thinking and various types of knowledge (episodic memory,

[1] In literature, general knowledge of an event (i.e., schema/script) and semantic memory of an event are sometimes used as synonymously and/or the latter is the source of the former (Berntsen & Bohn, 2010; Schacter et al., 2008).

autobiographic memory, schema-like representations, & semantic memory) because these are closely relevant to each other. First, an ability to self-generate a sequence of predictions is acquired as a consequence of daily unconscious activities. The external world continuously provides an event, and one implicitly predicts what follows next on the basis of what has happened so far (i.e., the past, especially recent episodes). Thus, our sequential prediction ability is based on our episodic memory (/autobiographic memory). Secondly, in a parallel-distributed processing (PDP) framework, general knowledge about a location/ context (schema/script) comes out as an emergent property, not built-in, through the act of sequential predictions of event sequences (Botvinick & Plaut, 2004; Schapiro et al., 2013; Rumelhart, et al., 1986). Specifically, a system does not need to access a stored "thing" or an isolable database about general knowledge of a location/context when it is interpreting the environment in order to predict what would come next. Rather, such a scheme-like behavior emerges in a system only by adjusting the connection strength among neuron-like processing units so that the system's function is tailored to the statistical structure of the event sequences. Finally, one may argue that such emergent knowledge underpinning sequential prediction (i.e., knowledge about what is likely to come next) is part of semantic knowledge about the ongoing event sequence.

Taken together, one may not need to build-in separable knowledge structures to simulate episodic future thinking. Instead, this study aimed (1) to train a model on the sequential prediction of *external event sequences* (i.e., the model receives an external event input, and is asked to predict the next external event input), and (2) to then allow the trained model to use this acquired knowledge for the *self-generation of internal event predictions* to simulate episodic future thinking (i.e., the model predicts the next event upon a event cue, and this prediction then forms the input of the next trial which in turn will trigger the next prediction). The specific predictions are as follows. After training, the nature of the acquired knowledge for external event predictions should reflect the characteristics of episodic memory, schema-like representations, and semantic memory. Thus, (A) the model should show a higher level of familiarity on the recently experienced/trained events than the remote episodes (i.e., episodic memory). Also, during sequential predictions of external events, (B) the model should be able to 'interpret' the context (e.g., now dining) of the current sequences (i.e., schema-like behavior). Related to this, one would say knowledge about what is likely to happen after a given event is part of semantic memory of that event. Thus, (C) the acquired representations (i.e., hidden layer activations) should mirror the structure of semantic memory, such that semantically-similar events should be represented more similarly. After demonstrating the nature of the knowledge acquired for sequential predictions of external events, the model was allowed to use this knowledge for self-generation of internal event predictions (its own output is the next input). If knowledge

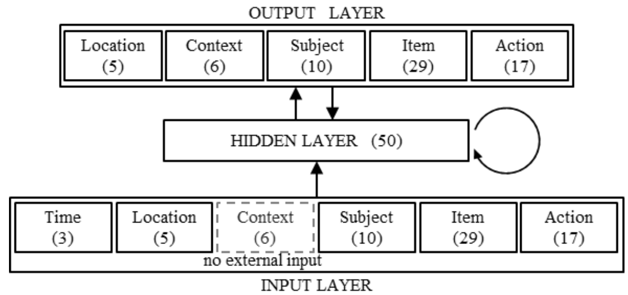

Figure 1: Three-layer simple-recurrent network. The number of units are shown in parentheses.

for sequential prediction of external events underlies episodic future thinking, then the model's self-generation of internal event predictions should capture the episodic future thinking of healthy participants. Thus, (D) the trained model should be able to self-generate a stable event sequencein the cued context. In contrast, if the computation of hidden layer activations is distorted (virtual 'lesioning'), then (E) semantically-similar items would not be represented in a similar manner, thereby mimicking the degraded conceptual knowledge seen in SD patients. (F) Such a lesioned model should find it difficult to self-generate event sequences in a specific context as real SD patients do. Importantly, like SD patients, a lesioned model should still show (G) a higher level of familiarity on recent episodes.

Method

Model Architecture, Task and Representations

Figure 1 shows the architecture of the model. The units in the two peripheral layers (input/output layers) were fully connected via units in the hidden layer in a feedforward manner. The activities in the hidden/output layers were fed back to the hidden layer at the next event through the (self-) recurrent connections. These recurrent connections enabled the model to gain 'memory' about past sequences (see Botvinick & Plaut, 2004; Elman, 1990). The input layer was divided into six sub-layers to represent each one of six elements of an event in a localist manner (see Table 1). For example, the units in each layer denoted: Time = [time1, time2, time3], Location = [home, school, university, office, town], Context = [dining, cooking, studying, working, cleaning, watching films], Item = [glass, knife, …etc.], Action = [take/grasp, place, bring into mouth, …etc.]. Given that the task was to predict the next input (e.g., Elman, 1990), the output layer was organized in the same way, except for the absence of the Time layer (in reality, we do not predict what the next Time is). Note, importantly, we assume that context label is not what is explicitly given from the external world (Rumelhart, et al., 1986). Therefore, units in the Context input layer did not receive any external input (these are written here so that the creation of the event sequences is clearer). This means that the model was never explicitly informed of the context label. On the other hand,

Table 1: Examples of the training sequences generated by cellular automata.

Event no.	Input (target of the previous trial) localist patterns						Events in English
	Time	Location	Context*	Subject	Item	Action	
1		school	dining	I	fork	take	I take a fork during dining at school.
2		school	dining	I	fork	stick	I stick a fork during dining at school.
3		school	dining	I	fork	bring to mouth	I bring a fork to my mouth during dining at school.
4		school	dining	I	fork	place**	I place a fork during dining at school.
5		school	dining	friend A	glass	take	A' takes a glass during dining at school.
6		school	dining	friend A	glass	bring to mouth	A' brings a glass to his mouth during dining at school.
7		school	dining	friend A	glass	place**	A' places a glass during dining at school.
8		school	cleaning	I	cup	wash	I wash a glass during cleaning at school.
9	chilhood (Time1)	school	cleaning	I	cup	place**	I place a glass during cleaning at school.
10		school	cleaning	friend B	towel	wipe	B' wipes with a towel during cleaning at school.
⋮		⋮	⋮	⋮	⋮	⋮	⋮
i		school	cleaning	I	cup	place**	I place a pen during studying at school.
i + 1		home	cooking	I	oil	open	I open oil during cooking at home.
i + 2		home	cooking	I	oil	pour	I pour oil during cooking at home.
i + 3		home	cooking	I	oil	place**	I place oil during cooking at home.
i + 4		home	cooking	I	knife	cut	I cut with a knife during cooking at home.
⋮		⋮	⋮	⋮	⋮	⋮	⋮
50,000,001	adolescence	college	studying	classmate F	pen	write	F' writes with a pen during studying at college.
⋮	(Time2)	⋮	⋮	⋮	⋮	⋮	⋮
100,000,001	adulthood	office	working	colleague H	PC	type	H' types texts into the PC at the office
⋮	(Time3)	⋮	⋮	⋮	⋮	⋮	⋮

Note . * Context input was not given to the network, but the output layer was required to switch 'ON' the correct Context unit.
Note . ** Once all the action lists were used up for each item, Action 'place' was taken. Then, from the next trial, Item information changed, and simultaneously, Location/Context/Subject information changed/unchanged probabilistically (see main text).

the units in the Context output layer received a target signal (i.e., the context label of the next input). This means that the model was required to interpret the context of the current event based on the other pieces of information available. Finally, a small amount of Gaussian noise (range = 0.2) was added to the input activations to reflect sampling variability.

Structure of Event Sequence

The event sequences in the real world are not random but follow certain statistical constraints. For example, a waiter is unlikely to pour wine before opening the bottle at a restaurant. Also, we are likely to bring a fork to our mouth after sticking it into food. Then, we place the fork down and grasp the glass of water, and so on. The event sequences for training were generated by cellular automata with similar statistical constraints to those within the real world. Table 1 shows the examples. The following statistical constraints were applied to the training sequence. Suppose the first event occurred (Event 1, in Table 1). Then, the following several events (Events 2-3) were generated with the same Time/Location/Context(not-presented)/Subject/Item information but with different Action information (randomly selected from the plausible Action lists for that Item, see Table 3). Once every possible action was selected without replacement, then 'place' Action was taken (see Event 4), at which point Item information of the next trial always (100%) changed, and simultaneously the other pieces of information changed probabilistically. Specifically, Subject information changed with a probability of 70%. And, if Subject changed, Context information changed with a probability of 30%. Then, if Context changed, Location information changed with a probability of 30%. The double asterisks in Table 1 denote examples of the timing when these probabilistic changes of information were made. When the cellular automata decided to change each piece of information, then the next piece of information was randomly selected from the possible lists (see Tables 2, 3) such that the event sequence was as realistic as possible. For example, Cooking context never occurred in Town; a frying-pan never appeared during Working, etc. Additionally, once an Item changed to another, the same Item was never selected until two other Items were selected. Finally, in order to simulate the recency effect on episodic memory (see Prediction A, above), the whole training sequence (300 million trials) was divided into three: The first 50 million trials were labeled as Time1 (Childhood), the next 50 million trials were Time2 (Adolescence), and finally the last 200 million trials were Time3 (Adulthood). As shown in Table 2, there was time-specific Location/Subject information (e.g., 'school' was specific in Time1). The input patterns with such time-specific information were later used to probe the model's episodic memory (see Figure 2 in Results and Discussion).

Training

In each trial, 64 units in the input layer were hard-clamped to their input values, and the activation spread in a feedforward manner. The error derivative was calculated (cross-entropy), and the weight was adjusted (back-propagation). A weight decay of 1E-9 was set. A learning rate of 0.41 was set, and was gradually reduced to 0.01 by 0.02 per every 10 million trials. An error derivative was set to zero if the target-output difference was less than 0.1.

Table 2: Possible Contexts/Subjects in each Time/Location

Location	Time	Context lists	Subject lists (and Time each Subject appears)
home	1-3	cleaning, studying, dining, cooking, watching films	I
school	1	cleaning, studying, dining	I, friend A (Time 1), friend B (Time 1), friend C* (Time 1)
college	2	cleaning, studying, dining	I, friend A (Time 2), classmate D (Time 2), classmate E (Time2), classmate F* (Time2)
office	3	cleaning, working, dining	I, classmate D (Time 3), colleague G (Time3), colleague H* (Time3)
town	1-3	dining, watching films	I, friend A (Time 1, 2, & 3), friend B (Time 1), classmate D (Time 2 & 3), classmate E (Time 2), colleague G (Time 3)

Note. * Friend C (Time 1), classmate F (Time 2), & colleague H (Time 3) are both time-/location-specific Subjects. Thus, any events with these Subjects were time-specific, which were used to probe the episodic memory (see main text).

Results and Discussion

Accuracy in Sequential Prediction Task

The task was not deterministic but probabilistic, such that the model was not able to predict the next event with absolute certainty. Elman (1990) evaluated the performance of such a task in terms of the cosine of the angle between the output vector and the target likelihood vector. The latter referred to the probability of each output unit to receive a target signal of 1.0 for a given input pattern. We were able to determine these probabilities from the training corpus, resulting in a mean cosine value collapsed over all the trials of 0.95 (*SD* = 0.01). Thus, the model successfully acquired the statistical structure underlining the event sequences.

Familiarity in Past Episodes (Prediction A)

Knowledge for sequential prediction of events should be acquired on the basis of past episodes (episodic memory). If the model has acquired episodic memory, then recently experienced patterns (Time3) should be more familiar than remote ones (Time1). Plaut (1997) measured the unit's polarity for an input pattern and took it as an index of that input pattern's familiarity. Figure 2 the distribution of the polarity values (taken from the output layer) for the time-specific trials (see Method). The polarity distribution of the most recent events (Time3) was higher than those of Time1/Time2, showing a recency effect.

Schema-like behavior (Prediction B)

The present model did not receive a context label (input) during training, but was trained to interpret the context of the current event sequences from the other pieces of

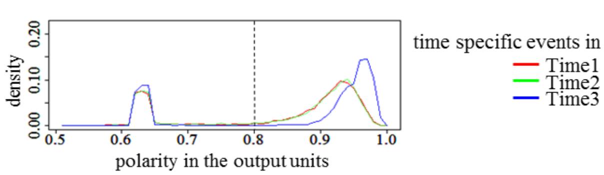

Figure 2: Distribution of the polarity values in the output layer units for time-specific events.

Table 3: Examples of possible Items/Action in each Context

Context	Item lists (and Action lists in each Item)
cleaning	glass (take/grasp, wash*, place), frying-pan (take/grasp, wash*, place), handkerchief (take/grasp, wash*, place), etc.
dining	glass (take/grasp, bring to a mouth*, place), handkerchief (take/grasp, wipe mouth*, place), etc.
cooking	frying-pan (take/grasp, shake/toss*, place), knife (take/grasp, cut, place) etc.
studying	pen (take/grasp, write, place), note (take/grasp, write, flip, look, place), etc.
working	handout/document (take/grasp, write, flip, read, place), PC (type, read) etc.
watching films	TV (watch), cinema screen (watch), popcorn (take/grasp, bring to mouth, place), etc.

Note . Full lists of Items and associated Actions in each Context/Location are available from authors upon request.
Note . * These actions are the examples of Context-specific actions (i.e., one does not wash handkerchief during dining).

information available (e.g., Item, Action). Such a schema-like behavior can be visualized by multi-dimensional scaling analysis on the hidden layer activation patterns (Schapiro et al., 2013). As shown in Figure 3, the trials during the same context were clustered together and separated from other trials (different context). Thus, as found with previous PDP models, general knowledge to interpret the outer world (i.e., schema) does not need to be built-in, but rather comes out as an emergent property (Rumelhart et al., 1986).

Emergence of Semantic Memory (Prediction C)

Acquisition of knowledge for sequential prediction means that the model knows what is likely to come next in a certain event. One would say such knowledge is part of semantic memory of an event. If so, semantically similar Items shouldbe similarly represented in the hidden layer. During training, each Item was presented with some realistic constraints (e.g.,glass appeared only in Cooking/Cleaning, and its plausible Action lists were take/grasp, stick, wash, etc. See Table 3). Therefore, if the model acquired semantic knowledge as an emergent property of sequential predictions, then semantically similar items in the real world (glass, cup, etc.) should be represented as similar patterns.

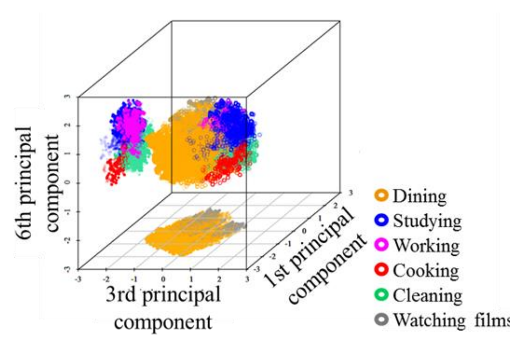

Figure 3: Multidimensional scaling 3D plot of the hidden layer activations underlying the model's context 'interpretation', and 2D-projections to improve visualization. The colors of the plots denote the input context label (not given to the network).

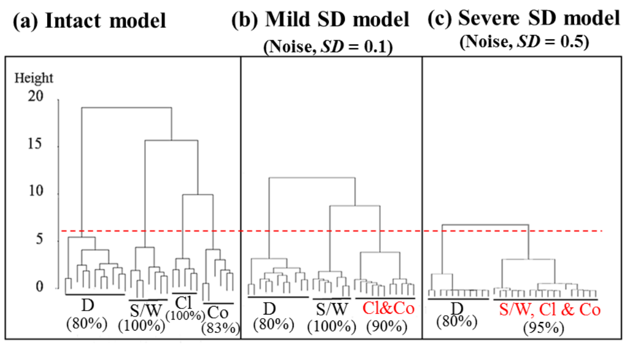

Figure 4: Cluster analysis of the hidden layer activations for each Item. Abbreviations used for labelling the clusters: D = dining, S/W = studying/working, Cl = cleaning, Co = Cooking. Percentages in parentheses denote the ratio of the Items that are correctly classified into four categories.

The hidden layer activation vectors for all the trials of an Item were averaged (collapsing over the other pieces of information) to form the representative vector of that Item, and we conducted a cluster analysis on those representative vectors of 29 Items. The left third of Figure 4 (intact model) shows the resultant dendrogram. If we draw aclustering criterion line as shown in Figure 4a (red holizontal line), then the data were represented by four clusters, and each cluster was interpreted to represent Items for Dining, Studying/Working, Cleaning, and Cooking, respectively. Thus, the hidden layer activities captured the semantic similarities of Items in the real world, which was a signature of emergent semantic memory.

Episodic Future Thinking (Prediction D)

The analyses so far revealed that the model acquired knowledge for sequential prediction of external events, and this knowledge captured the nature of episodic memory, schema-like representations, and semantic memory of Items. Next, we tested if this knowledge underlied simulation of future events. First, the model was presented with an event pattern (e.g. Colleague H brought a cup to his mouth during *dining* at the office) as a cue just like in a human experiment. Once the model predicted the next event, this output vector was converted to binary in a 'winner-takes-all' manner,. and was hard-clamped to the input layer at the next trial (e.g., if Glass output unit had the highest activation value among the Item units, then the corresponding input unit was clamped to 1.0 at the next trial). This cycle was repeated 300 times. Figure 5a (Intact model) plotted the Context output of the self-generated internal predictions (a holizontal line means a stable internal prediction sequence within the same context). After training, the self-generated sequence in the intact model was stable in the same context as the cued information, mirroring the episodic future thinking of normal participants (Irish et al, 2012).

Simulating Semantic Dementia

SD patients' future simulations tended to deviate from the cued context (Irish et al., 2012). If knowledge for sequential prediction underlies episodic future thinking, then our model should simulate an impaired behavior as well. To simulate different levels of severity, different levels (*SD* = 0.01, 0.1, & 0.5, respectively) of noise were added to the recurrent connection (e.g., Botvinick & Plaut, 2004).

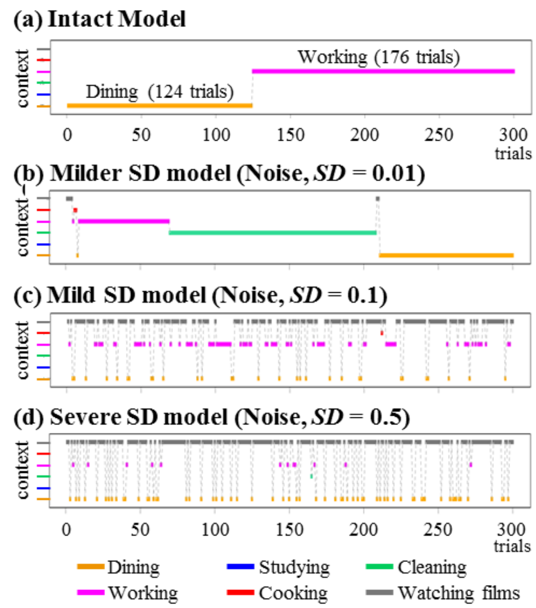

Figure 5: Survival plots of the Context output during internal predictions against the number of predictions by (a) Intact, (b) milder, (c) mild, & (d) severe SD models.

Degradation of Semantic Memory (Prediction E) SD patients exhibit more degraded semantic memory as the disease progresses. Figure 4b and 4c shows the results of the same cluster analysis as before on the data from the damaged models. Although there were four semantic categories to classify Items in the intact model (Figure 4a) such categories got more blurry as the amount of the noise increased. Moreover, the distance (y-axis height) between items within a category also got shorter. This means the semantically-similar items became undiscriminable from each other. Thus, semantic memory of the damaged model was degraded as observed in real patients (Irish et al., 2012).

Episodic Future Thinking (Prediction F) Next, the damaged models were tested on their ability to self-generate a sequence of internal event predictions. Three levels of severity were simulated (Figures 5b, 5c, and 5d). As the noise level increased, context information in the self-generated sequence shifted more frequently from the cued event (dining). Thus, like real SD patients, the damaged model could not maintain the internal event predictions in a specific context. Moreover, context information generated was incongruent with location (e.g., watching films in the Office, which never occurred during training).

Relatively Intact Episodic Memory (Prediction G) Figure 6a and 6b show the impact of lesioning on the distribution of the polarity (familiarity) values for the time-specific

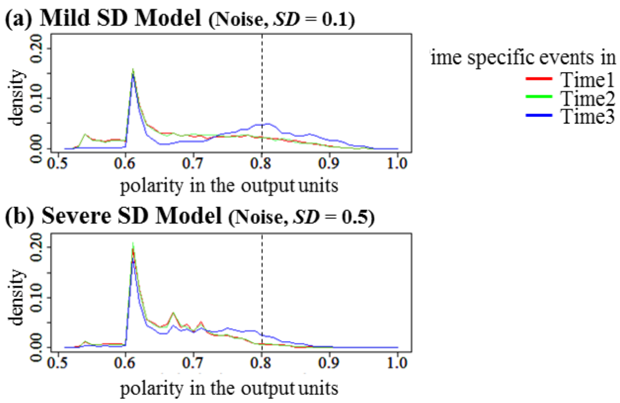

Figure 6: Distribution of the polarity values in the output layer units for time-specific events of damaged models

events. Like real SD patients (Irish et al., 2011), the familiarity (polarity) of the recent events (Time3) remained high when the noise range was small (mild SD). However, such a recency effect disappeared for the severe model (Figure 6b). Taken together, these patterns replicated the dissociation between degraded semantic memory and relatively intact episodic memory (especially for recent events) found in real SD patients (Irish et al., 2011). Thus, impaired episodic future thinking in SD patients can be attributed to degradation of semantic memory.

Summary and Conclusion

Episodic future thinking involves the successive self-generation of internal event predictions. The present model acquired knowledge that captured the characteristics of episodic memory, schema-like representation (to interpret context) and semantic memory through sequential prediction of external events. The model used such knowledge to conduct a self-generation of internal event predictions. The resultant sequences mirrored the episodic future thinking of healthy people. In contrast, a damaged model mimicked the profile of SD patients, and it had a difficulty in self-generating internal event predictions in a steady context. These findings suggest that semantic memory contributes to episodic future thinking so that temporally extended event sequences settle into a steady state within a cued context. An isolable knowledge structure (e.g., schema) does not need to be built-in, but rather such knowledge emerges from daily sequential prediction. Another implication from this model is that event and object knowlege do not require different subsystems, but could be represented in a multidimensional space of a single hidden layer. This is consitstent with Schapiro et al. (2013) who argued the similar computational principles for object semantics and event knowledge. Note before that we do not mean a one-to-one correspondence between event knowledge and episodic memory, and therefore this does not mean a single-system account for episodic memory and semantic memory, neither. Further insight on this issue would be gleaned by extending this model into other episodic/semantic tasks.

Finally, a tempting idea is that episodic future thinking can screen different types of neulogical patients given the relationship between semantics and context-coherent event predictions we demonstrated. At this point, we should be cautious because it requires a detailed error analysis, but this would be an intial step towards such a clinical contribution.

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.

References

Botvinick, M., & Plaut, D. C. (2004). Doing without schema hierarchies: a recurrent connectionist approach to normal and impaired routine sequential action. *Psychological review, 111*(2), 395–429.

Berntsen, D., & Bohn, A. (2010). Remembering and forecasting: The relation between autobiographical memory and episodic future thinking. *Memory & cognition*, *38*(3), 265–278.

D'Argembeau, A., & Mathy, A. (2011). Tracking the construction of episodic future thoughts. *Journal of experimental psychology: General*, *140*(2), 258–271.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(5), 1726–1731.

Irish, M., Hornberger, M., Lah, S., Miller, L., Pengas, G., Nestor, P. J., Hodges, J. R., Piguet, O. (2011). Profiles of recent autobiographical memory retrieval in semantic dementia, behavioural-variant frontotemporal dementia, and Alzheimer's disease. *Neuropsychologia, 49*(9), 2694–2702.

Irish, M., Addis, D. R., Hodges, J. R., & Piguet, O. (2012). Exploring the content and quality of episodic future simulations in semantic dementia. *Neuropsychologia*, *50*(14), 3488–3495.

Plaut, D. C. (1997). Structure and Function in the Lexical System: Insights from Distributed Models of Word Reading and Lexical Decision. *Language and Cognitive Processes*, *12*(5/6), 765–806.

Rumelhart, D. E., & McClelland J. L., & the PDP Research Group (1986) Parallel distributed processing: Explorations in the microstructure of cognition (Vol. 2). Cambridge, MA: MIT Press

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. (2013). Neural representations of events arise from temporal community structure. *Nature neuroscience*, *16*(4), 486–492.