Improving Neural Models of Language with Input-Output Tensor Contexts

Eduardo Mizraji[1(✉)], Andrés Pomi[1], and Juan Lin[1,2]

[1] Group of Cognitive Systems Modeling, Biophysics Section, Facultad de Ciencias, Universidad de la República, Iguá 4225, 11400 Montevideo, Uruguay
emizraji@gmail.com, andres.pomi@gmail.com, jlin2@washcoll.edu

[2] Department of Physics, Washington College, Chestertown, MD 21620, USA

Abstract. Tensor contexts enlarge the performances and computational powers of many neural models of language by generating a double filtering of incoming data. Applied to the linguistic domain, its implementation enables a very efficient disambiguation of polysemous and homonymous words. For the neurocomputational modeling of language, the simultaneous tensor contextualization of inputs and outputs inserts into the models strategic passwords that rout words towards key natural targets, thus allowing for the creation of meaningful phrases. In this work, we present the formal properties of these models and describe possible ways to use contexts to represent plausible neural organizations of sequences of words. We include an illustration of how these contexts generate topographic or thematic organization of data. Finally, we show that double contextualization opens promising ways to explore the neural coding of episodes, one of the most challenging problems of neural computation.

Keywords: Matrix memories · Tensor contexts · Word strings
Semantic spaces · Episodic memory

Gradually, it saw itself (like us)
imprisoned in this sonorous web
of Before, After, Yesterday, While, Now,
Right, Left, Me, You, Those, Others.

From "The Golem" by J.L. Borges

1 Introduction

The procedures developed by the human brain to organize sequences of semantic elements that create meaningful phrases are yet an unsolved problem. Such a sequence can be metaphorically congruent to the search for the exit of an intricate labyrinth, with myriad galleries connecting thousands of semantic modules. In this labyrinth, the output of a module is specifically guided toward its next module, a process that generates a completely non-random sequence of words. This controlled guidance can be due to the existence of specific "keys" that select and open the next appropriate semantic target.

A. Karpov et al. (Eds.): SPECOM 2018, LNAI 11096, pp. 430–440, 2018.
https://doi.org/10.1007/978-3-319-99579-3_45

Taking into account the extremely large number of possibilities offered by the semantic network, the possibility of building rapid meaningful phrases in natural language strongly suggests that these output keys explore all their potential targets in parallel.

An interesting approach would be to consider the creation of a meaningful phrase as analogous to the production of a sequence of motor acts oriented toward a goal [1–5]. This analogy would assume that before the construction of a phrase there exists an objective that induces a layout over which the words are organized. In this case the goal is a communicational task, and a complete discourse can be structured by a set of sub-targets that organize their parts.

In this work we shall try to model the emergence of different kinds of language organization, by representing semantic modules with matrix associative memories. The many remarkable properties of these matrix memories are described in [6–10]. As "mesoscopic models" they connect algorithms operating on complex symbolic data to the neuro-dynamic level [11]. In this formalism, to find a path in the labyrinth of semantic modules would mean that outputs of matrix associative memories become inputs of particular memories that produce the words in the general layout of the phrase that is being created. Our contribution aims to fill this framework by showing that the modulation of inputs and outputs of matrix memories by tensor contexts provides a procedure to explain how coherent sequences of words can be created. In addition, this formalism implies the possibility of building thematic clusters in semantic spaces.

2 Basic Models

In what follows we describe some properties of matrix associative memories and how tensor contexts enlarge their computational abilities.

2.1 Matrix Associative Memories

A matrix memory associates an m-dimensional column input vector f_i to an n-dimensional output vector g_i. Kohonen [10] shows that a memory can be characterized by the set

$$\mathrm{Mem} = \left\{(g_1, f_1), (g_2, f_2), \ldots, \left(g_Q, f_Q\right)\right\}. \tag{1}$$

This "learning set" represents the data to be stored in a matrix memory M. To find the appropriate structure of this matrix, define two partitioned matrices

$$G = \left[g_1\, g_2\, \cdots g_Q\right],\;\; F = [f_1\, f_2\, \cdots f_Q],$$

and represent the associations between the Q pairs of vector patterns by the matrix equation $G = MF$. Let $\mathrm{In} = \{1, 2, \ldots, Q\}$ be the set of indexes of stored pairs. Under this condition, the best solution in the sense of least squares, is in terms of the pseudoinverse F^+:

$$M = GF^+. \tag{2}$$

In the extremely simple case of an orthonormal set of inputs $\{f_i\}$, $i = 1$ to Q, Eq. (2) admits the closed expression:

$$M = \sum_{i=1}^{Q} g_i f_i^T. \tag{3}$$

For this matrix memory the recall operates as follows:

$$Mf_k = \sum_{i=1}^{Q} g_i \langle f_i, f_k \rangle, \tag{4}$$

with the scalar product being $\langle f_i, f_k \rangle = \delta_{ik}$ (δ_{ik} is the Kronecker's delta); hence if the index $k \in$ In, the recall is perfect, $Mf_k = g_k$.

2.2 Input Tensor Contexts

Imagine we need to model a neural network capable to disambiguate homonymic or polysemic words. Networks with hidden layers trained with backpropagation, are the classical devices to deal with this kind of problem [12]. However, in such approach we generally lose the possibility of a transparent mathematical theory allowing to predict what is happening during training as well as the final network structure. This opacity was the main motivation to develop a "transparent connectionist" alternative [13]. This alternative uses a kind of vector symbolic architecture based on tensor contextualization [11, 14, 15].

Let f_i be one homonymic word, associated with two vectors g_{i1} and g_{i2} for two completely non-correlated concepts. For instance, the input can represent the word "bank" and one output would be "money" and the other would be "sand". To retain the matrix format of the associative memory, we integrate the input with two vector contexts $p_{i1}, p_{i2} \in \mathbb{R}^h$ using the Kronecker product $\otimes$, a tensor procedure adapted to the operations of matrix algebra [16]. In our example, we could consider that the first context concerns finances and the second geography. The segment of a memory in our example can be expressed as:

$$M_i = g_{i1}(p_{i1} \otimes f_i)^T + g_{i2}(p_{i2} \otimes f_i)^T. \tag{5}$$

Consequently, when the memory receives an input and the corresponding context, the selection of the output happens via two scalar products:

$$M_i(p_{12} \otimes f_i) = g_{i1}\langle p_{i1}, p_{i2} \rangle \langle f_i, f_i \rangle + g_{i2}\langle p_{i2}, p_{i2} \rangle \langle f_i, f_i \rangle. \tag{6}$$

In a situation where both, the inputs and the contexts are orthonormal, we have a resolution of ambiguity,

$$M_i(p_{12} \otimes f_i) = g_{i2}. \tag{7}$$

This format can be generalized [14, 17, 18] to a global memory module composed of a variety of specialized sub-modules, each having the required complexity for the contextualization of its inputs:

$$M = \sum_i M_i. \tag{8}$$

2.3 Input-Output Contexts

We can extend the previous approach by modulating both, inputs and outputs with vector contexts. This approach leads to memory matrices with the following general structure:

$$H = \sum_{i,j,k} \left(p'_{ik} \otimes g_{ij}\right)\left(p_{ik} \otimes f_{ij}\right)^T. \tag{9}$$

From the properties of Kronecker products, the H matrix admits some interesting alternative representations. We illustrate two of them:

$$H = \sum_{i,j,k} (p'_{ik}\, p^T_{ik}) \otimes \left(g_{ij} f^T_{ij}\right), \tag{10}$$

$$H = \sum_{i,j,k} \left(p'_{ik} \otimes I_{\dim(g_{ij})}\right) g_{ij} f^T_{ij} \left(p_{ik} \otimes I_{\dim(f_{ij})}\right)^T. \tag{11}$$

Note that inputs $p_{uv} \otimes f_{ab}$ with stored patterns, display outputs given by

$$H(p_{uv} \otimes f_{ab}) = p'_{uv} \otimes g_{ab}. \tag{12}$$

These outputs are prepared to enter as inputs to a similar memory H' with this particular pair [context - pattern] stored in its database.

Memories with this structure accept many representational and computational potentialities to process the operations displayed by natural languages [19, 20]. In the next Sections we shall describe some of these operations.

3 Deterministic Semantic Strings

In his "Principles of Psychology" (Vol. II, Chap. XXVI) James [21] writes that voluntary acts are based on consolidated memory traces created by previous involuntary acts. Similarly, the voluntary creation of phrases has as prerequisite the existence of word associations in previously fixed memories–developed after experiential contact with word usage.

As we mentioned before language production could be seen as the generation of meaningful phrases, and may be similar to the assembly of a sequence of motor actions aimed at reaching a goal [3, 4, 22]. The purpose of spoken or written phrases is to transmit information by means of expressions that can be understood. Neural modeling challenges us to reach this goal by triggering an appropriate chain of meaningful words.

Let us suppose that a phrase could be represented by a string:

$$F(\alpha, n) = \langle a_{\alpha 1},\, a_{\alpha 2}, \ldots, a_{\alpha n} \rangle,\; a_{\alpha i} \in \mathrm{Sem}\{a_1,\, a_2, \ldots, a_\omega\}, \tag{13}$$

with Sem being the very large set of words in a normal lexicon. The phrase can repeat words, and consequently it is possible to have $a_{\alpha i} = a_{\alpha j}$. Now, how do we insure that $a_{\alpha 1}$ precedes $a_{\alpha 2}$? Moreover, how does the meaning of the phrase guide the correct order of successive words while information is transmitted? A possible answer to the first question would be to assume that the transition probabilities between words are responsible for the correct sequence, with a given word followed by its most probable successor. Within this framework, language production is mainly represented by a stochastic process with transition probabilities dependent on segments of previously used words. [23–26]. The second question seems to imply the existence of an anticipatory layout for the phrase.

Here, we explore the following proposal. Imagine a small string of three words $\langle a_{\alpha 1},\, a_{\alpha 2},\, a_{\alpha 3} \rangle$ representing a miniature phrase. Let us immerse these elements in contexts, generating a new string

$$\left\langle \left(p_{targ}\, a_{\alpha 1}\, p_1\right),\, \left(p_1\, a_{\alpha 2}\, p_2\right),\, \left(p_2\, a_{\alpha 3}\, p_{end}\right) \right\rangle, \tag{14}$$

The neural vector p_{targ} is both, the context that triggers the sequence and concurrently, the target code. Contexts p_1 and p_2 are keys indicating the correct next element of the string, and context p_{end} marks the end of the phrase. In this way, a good sequence of words is selected by the contextual string

$$\left\langle p_{targ},\, p_1,\, p_2,\, p_{end} \right\rangle. \tag{15}$$

A recursive tensor input-output memory with the structure

$$S = (p_{end} \otimes a_{\alpha 3})(p_2 \otimes a_{\alpha 2})^T + (p_2 \otimes a_{\alpha 2})(p_1 \otimes a_{\alpha 1})^T + (p_1 \otimes a_{\alpha 1})p_{targ}^T \tag{16}$$

can accomplish the procedure just described. In the general case, the final output can be a "pure" string of words, $\langle a_{\alpha 1}, a_{\alpha 2}, \ldots, a_{\alpha n} \rangle$. The contexts used in an internal, hidden computation, are channeled by a filter Way Out Matrix (WOM) having the structure

$$\mathrm{WOM} = \left(\sum_\lambda p_\lambda^T \right) \otimes I_{\dim(a)}. \tag{17}$$

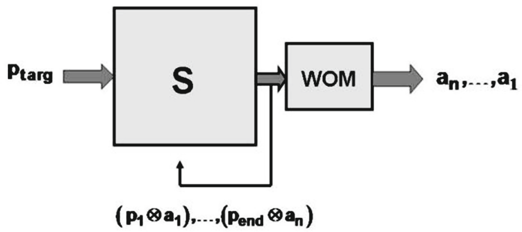

Fig. 1. This diagram illustrates how a context target enters a recursive semantic network S triggering a sequence of contextualized outputs. These outputs are filtered by a WOM matrix that extracts the contexts and produces a pure word string.

The sum includes all the relevant contexts, and $I_{dim(a)}$ is an identity matrix with the same dimension as word vectors. Note that

$$\mathrm{WOM}\left(p_\gamma \otimes a_h\right) = a_h. \tag{18}$$

In Fig. 1 we illustrate this recursive model for a string of arbitrary length.

The neurobiology of lexical strings production is far from being understood. We can consider the voluntary construction of utterances by our model in light of William James' thought. Our model requires the previous existence of permanent memories of words and contextual markers, and a transitory working memory to install the appropriate string format. Finally, we mention that the target 'feeds and builds' contexts to generate meaningful strings in the same way that the target of a mechanical movement of our arm guides the intermediate steps needed to reach it.

4 Clustering by Contexts

The memory H given in Eq. (10), with sets of different input-output associations sharing the same pair of input-output contexts can be factorized into clusters of associations induced by the contexts,

$$H = \sum_i \left[p'_i\, p_i^T \otimes \sum_j g_{ij}\, f_{ij}^T \right]. \tag{19}$$

This partition suggests how scattered data may be organized in large neural networks. Contexts may create a topical coherence in a recall. Let us mention that an interesting formal parallelism between matrix memories and the Latent Semantic Analysis (LSA) has been described in [19]. In this direction, the structure of matrices (10) and (19) suggests the possibility of looking for the thematic clustering of text-document matrices using, instead of a classical LSA based on SVD, a procedure that labels topics via the search of Kronecker factors.

If we use as contexts unit vectors e_s (vectors with a 1 in position s and 0's otherwise), the matrix H can be expressed as:

$$H = \sum_i \left[e_i' e_i^T \otimes M_{ij}\right], \tag{20}$$

with

$$M_{ij} = \sum g_{ij} f_{ij}^T \tag{21}$$

being a classical Anderson-Kohonen associative memory matrix. By an adequate selection of dimensions for the context unit vectors, it is possible to generate a topographic pattern with different associative memories M placed as tiles into the "host" matrix H (Pomi, Mizraji and Lin, paper submitted). We illustrate this point with a simple example. Given the two unit column vectors

$$e_1 = [1 \quad 0]^T; \; e_2 = [0 \quad 1]^T$$

and four associative memory matrices, $M^{(v)} \in \mathbb{R}^{p \times q}$, $v = 1, \ldots, 4$

H takes the form

$$H = e_1 e_1^T \otimes M^{(1)} + e_1 e_2^T \otimes M^{(2)} + e_2 e_1^T \otimes M^{(3)} + e_2 e_2^T \otimes M^{(4)}. \tag{22}$$

After computing the Kronecker products we find

$$H = \begin{bmatrix} M^{(1)} & M^{(2)} \\ M^{(3)} & M^{(4)} \end{bmatrix}, \; H \in \mathbb{R}^{2p \times 2q}. \tag{23}$$

Thus, the contexts create a computational layer composed by various memory modules located in specific topographies, each one able to receive and redirect information selectively channeled by the contexts.

Kohonen [27] developed one of the most important and deep procedures to model the generation of topographic neural patterns. The approach we are describing here assumes cognitive supervised learning. One could imagine associative memories to be the result of active interactions between a trainable brain and an external instructor–an active human teacher or environmental experiences. Hence, emergent clusters of associative memories may explain how, after extensive vocabulary learning, complex semantic webs can be established. We want to mention that the results of Huth et al. [28] experimentally illustrate the existence of a remarkable topographic organization in the semantic web of the human brain.

5 Episodes

Since the foundational characterization of episodic memories by Tulving (updated in [29]), the search for their neural bases became an important research objective [30–34]. Adapting ideas of these investigators, we shall assume that episodic memories result from the interaction of different classes of memories, fundamentally, a semantic memory and a context memory that stores episode markers. We illustrate the interaction between these memory modules in Fig. 2.

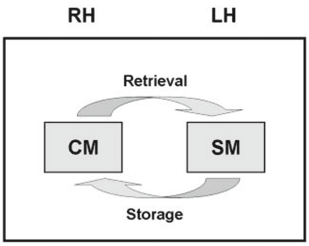

Fig. 2. This scheme adapts to our model one of the conceptions about episode storage and retrieval. LH: Left hemisphere, RH: Right hemisphere, SM: Semantic Memory, CM: Contexts Memory.

We are going to assume that the encoding happens mainly in a region capable of sustaining a semantic memory (e.g.: the left prefrontal cortex) and the recall involves a region that stores contextual markers (e.g.: the right prefrontal cortex). The model we want to comment is formally similar to the model that generates semantic strings. However, there is a crucial difference: in episodes we do not necessarily have a target. A contingent series of events is stored in the memory due to a variety of causes, among others, emotional impact, autobiographical importance, bizarre consequences, etc. In these episodic sequences, contexts provide a kind of positional information–an expression of the embryologist Lewis Wolpert–that places words in the precise positions needed to recreate the episode.

Let us define an episode by a time sequence of contexts that intermingle with words selected from the semantic memory. The sequence of contexts can be generated by a cyclic memory structured as:

$$C = p_{out}\, p_n^T + p_n\, p_{n-1}^T + \cdots + p_1\, p_{in}^T. \tag{24}$$

Context vector p_{in} marks the beginning of the sequence, and context p_{out} marks the end. Within a recursive network, the reinjection of successive outputs of memory C creates the time pattern

$$\langle p_{out},\, p_n, \ldots, p_i,\, p_{in} \rangle. \tag{25}$$

Intermingling these contexts with words a_i extracted from the semantic memory, builds the episodic sequence

$$\langle (p_{out}\, a_n\, p_n), (p_n\, a_{n-1}\, p_{n-1}), \ldots, (p_3\, a_2\, p_2)\,,\ (p_2\, a_1\, p_{in})\rangle. \tag{26}$$

We are going to model this situation by assuming that intermingling occurs because the semantic memory is structured with associative memories that can be approximated by matrices like

$$E = \sum_{i,j,k} \left(p'_{ik} \otimes a_{ij}\right)\left(p_{ik} \otimes a_{ij}\right)^T, \tag{27}$$

with the particularity that context markers are very sparse vectors (e.g.: unit vectors). The total set of stored episodes can be based on a semantic basis of N words, N being very large. A given memory cannot store all this variety due to dimensional limitations. But memories like (25) can surpass the dimensional limitations imposed by neuroanatomy and enlarge the variety of episodes via a multi-modular semantic organization. The final step of the episodic recall can be a pure verbal string emerging from a WOM filter.

We end this Section by mentioning that there is a close relationship between remembered episodes, and episodes created by the imagination. A fictional story does not travel to the autobiographical past, but creates episodes that we can recall even if such episodes are placed in the far past or future. This shows an interesting point concerning the possible coincidence between the neural systems responsible for the recall of personal biographical episodes and the imaginary generation of fictional facts (see [35, 36] for extensive references about this point), including the conception of innovative literary, philosophical, scientific, or technological scenarios.

6 Perspectives

In this work we have assumed that a semantic unit, integrated with many contexts, could participate in a large variety of different linguistic tasks. The described models are written in terms of matrix algebra and Kronecker tensor products, which makes them operationally transparent and easily amenable to computer implementation, even though the dimensions involved in these linguistic tasks can be extremely large. In any case, the highly flexible production of organized, non-random sequences of words in a natural language is a marvelous and yet obscure process. The topical organization of a biological semantic web, with patches including elaborate pieces of language could plausibly be a basis for the hierarchical elaboration of complex thoughts. These thoughts are translated into linguistic codes and communicated. In a way, "deep learning" technological procedures involving a system of hierarchical computing levels, are already implemented by the human brain. We need to understand these codes, which in many cases, can be accompanied by linguistic productions. A simplified example of this kind of hierarchical processing is given in [20]. Finally, the recreation, or invention of episodes represents one of the most significant signatures of the human mind and is placed, by researchers like Tulving [29], at the highest levels of cognition. With tensor input-output contexts we have been able to formulate an elementary approach to the modeling of these open and crucial problems.

Acknowledgments. AP and EM acknowledge partial financial support by PEDECIBA and CSIC-UdelaR.

References

1. Luria, A.R.: The Working Brain. Basic Books, New York City (1973)
2. Kimura, D.: Neuromotor mechanisms in the evolution of human communication. In: Steklis, H.D., Raleigh, M.J. (eds.) Neurobiology of Social Communication in Primates, pp. 197–219. Academic Press, New York (1979)
3. Calvin, W.H.: A stone's throw and its launch window: timing precision and its implications for language and hominid brains. J. Theor. Biol. **104**, 121–135 (1983)
4. Calvin, W.H.: The unitary hypothesis: a common neural circuitry for novel manipulations, language, plan-ahead, and throwing? In: Gibson, K.R., Ingold, T. (eds.) Tools, Language, and Cognition in Human Evolution, pp. 230–250. Cambridge University Press, Cambridge (1993)
5. Ojemann, G.A.: Brain organization for language from the perspective of electrical stimulation mapping. Behav. Brain Sci. **6**, 189–206 (1983)
6. Anderson, J.A.: A simple neural network generating an interactive memory. Math. Biosci. **14**, 197–220 (1972)
7. Anderson, J.A.: An introduction to neural networks. MIT Press, Cambridge (1995)
8. Cooper, L.N.: A possible organization of animal memory and learning. In: Lundquist, B., Lundquist, S. (eds.) Proceedings of the Nobel Symposium on Collective Properties of Physical Systems, pp. 252–264. Academic Press, New York (1973)
9. Kohonen, T.: Correlation matrix memories. IEEE Trans. Comput. **C-21**, 353–359 (1972)
10. Kohonen, T.: Associative Memory: A System Theoretical Approach. Springer, Heidelberg (1977). https://doi.org/10.1007/978-3-642-96384-1. Chap. 3
11. Beim Graben, P., Potthast, R.: Inverse problems in dynamic cognitive modeling. Chaos Interdiscip. J. Nonlinear Sci. **19**, 015103 (2009)
12. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**, 533–536 (1986)
13. Carmantini, G.S., Beim Graben, P., Desroches, M., Rodrigues, S.: A modular architecture for transparent computation in Recurrent Neural Networks. Neural Netw. **85**, 85–107 (2017)
14. Mizraji, E.: Context-dependent associations in linear distributed memories. Bull. Math. Biol. **51**, 195–205 (1989)
15. Smolensky, P.: Tensor product variable binding and the representation of symbolic structures in connectionist systems. Artif. Intell. **46**, 159–216 (1990)
16. Graham, A.: Kronecker Products and Matrix Calculus With Applications. Ellis Horwood, Chichester (1981)
17. Pomi, A., Mizraji, E.: Semantic graphs and associative memories. Phys. Rev. E **70**, 066136 (2004)
18. Pomi, A.: A possible neural representation of mathematical group structures. Bull. Math. Biol. **78**, 1847–1865 (2016)
19. Mizraji, E., Pomi, A., Valle-Lisboa, J.C.: Dynamic searching in the brain. Cogn. Neurodyn. **3**, 401–414 (2009)

20. Mizraji, E., Lin, J.: Modeling spatial-temporal operations with context-dependent associative memories. Cognit. Neurodyn. **9**, 523–534 (2015)
21. James, W.: Principles of Psychology. The Great Books of the Western World, vol. 53. The University of Chicago (1890)
22. Nishitani, N., Schürmann, M., Amunts, K., Har, R.: Broca's region: from action to language. Physiology **20**, 60–69 (2005)
23. Jurafsky, D., Bell, A., Gregory, M., Raymond, W.D.: Probabilistic relations between words: evidence from reduction in lexical production. Typol. Stud. Lang. **45**, 229–254 (2001)
24. Jurafsky, D.: Probabilistic modeling in psycholinguistics: linguistic comprehension and production. In: Bod, R., Hay, J., Jannedy, S. (eds.) Probabilistic Linguistics, p. 21. MIT Press, Cambridge (2003). Chap. 3
25. Nowak, M.A., Komarova, N.L., Niyogi, P.: Computational and evolutionary aspects of language. Nature **417**, 611–617 (2002)
26. Chater, N., Manning, C.D.: Probabilistic models of language processing and acquisition. Trends Cognit. Sci. **10**, 335–344 (2006)
27. Kohonen, T.: Self-Organizing Maps. Springer, Heidelberg (1997). https://doi.org/10.1007/978-3-642-97966-8
28. Huth, A.G., Nishimoto, S., Vu, A.T., Gallant, J.L.: A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron **76**, 1210–1224 (2012)
29. Tulving, E.: Episodic memory. Annu. Rev. Psychol. **53**, 1–25 (2002)
30. Baddeley, A.: Working memory: looking back and looking forward. Nat. Rev. Neurosci. **4**, 829–839 (2003)
31. Jonides, J.R., et al.: The mind and brain of short-term memory. Ann. Rev. Psychol. **59**, 193–224 (2008)
32. Repovs, G., Baddeley, A.: The multi-component model of working memory: explorations in experimental cognitive psychology. Neuroscience **139**, 5–21 (2006)
33. Eichenbaum, H.: Prefrontal–hippocampal interactions in episodic memory. Nature Rev. Neurosci. **18**, 547–558 (2017)
34. Schapiro, A.C., Turk-Browne, N.B., Botvinick, M.M., Norman, K.A.: Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. Philos. Trans. R. Soc. Lond B **372**, 20160049 (2017)
35. Schacter, D.L., et al.: The future of memory: remembering, imagining, and the brain. Neuron **76**, 677–694 (2012)
36. Schacter, D.L., Benoit, R.G., Szpunar, K.K.: Episodic future thinking: mechanisms and functions. Curr. Opin. Behav. Sci. **17**, 41–50 (2017)