Interactive Control of Computational Power in a Model of the Basal Ganglia-Thalamocortical Circuit by a Supervised Attractor-Based Learning Procedure

Jérémie Cabessa[1] and Alessandro E.P. Villa[2(✉)]

[1] Laboratoire d'économie Mathématique (LEMMA),
Université Paris 2 – Panthéon-Assas,
4, Rue Blaise Desgoffe, 75006 Paris, France
[2] NeuroHeuristic Research Group, University of Lausanne,
Quartier UNIL-Dorigny, 1015 Lausanne, Switzerland
avilla@unil.ch
http://www.neuroheuristic.org

Abstract. The attractor-based complexity of a Boolean neural network refers to its ability to discriminate among the possible input streams, by means of alternations between meaningful and spurious attractor dynamics. The higher the complexity, the greater the computational power of the network. The fine tuning of the interactivity – the network's feedback output combined with the current input stream – can maintain a high degree of complexity within stable domains of the parameters' space. In addition, the attractor-based complexity of the network is related to the degree of discrimination of specific input streams. We present a novel supervised attractor-based learning procedure aimed at achieving a maximal discriminability degree of a selected input stream. With a predefined target value of discriminability degree and in the absence of changes in the internal connectivity matrix of the network, the learning procedure updates solely the weights of the feedback projections. In a Boolean model of the basal ganglia-thalamocortical circuit, we show how the learning trajectories starting from different configurations can converge to final configurations associated with same high discriminability degree. We discuss the possibility that the limbic system may play the role of the interactive feedback to the network studied here.

Keywords: Boolean recurrent neural networks · Learning · Attractors · Plasticity · Interactivity · Basal ganglia-thalamocortical circuit · Limbic system

1 Introduction

Attractor dynamics or quasi-attractor dynamics have been associated to perceptions, thoughts and memories, and the chaotic itinerancy between those with

© Springer International Publishing AG 2017
A. Lintas et al. (Eds.): ICANN 2017, Part I, LNCS 10613, pp. 334–342, 2017.
https://doi.org/10.1007/978-3-319-68600-4_39

sequences in thinking, speaking and writing [1–3]. Specific spike trains – time series defined by the epochs of neuronal discharges – were experimentally shown to be associated with such (chaotic) attractor dynamics [4–8]. Moreover, experimental neurophysiological studies suggest that spatiotemporal patterns of discharges repeating more often than expected by chance may be associated to processing and coding of information in the brain, in particular in association with specific behaviors [9–14]. Spatiotemporal patterns have also been observed in simulations of nonlinear dynamical systems [15,16] as well as in simulations of large scale neuronal networks [17]. Hence, the correlation between attractor dynamics and recurrent spatiotemporal patterns of discharges has been suggested as an alternative model to synfire chains [3,9,18]. Neural coding of perceptual and contextual information may be performed by underlying attractor dynamics, with the advantage of implicit transmission and storage of memory traces in the network connectivity [19].

We introduced an *attractor-based complexity* measure for Boolean recurrent neural networks, which refers to the ability of the networks to discriminate among their possible input streams, by means of alternations between meaningful and spurious attractor dynamics [18,20,21]. This complexity measure is assumed to be related to some aspects of the computational capabilities of Boolean neural networks. It was applied to study the attractor dynamics of a Boolean model of the basal ganglia-thalamocortical circuit [18]. In this model, the attractor dynamics and its associated complexity measure are highly sensitive to local and global modifications of the coupling strength between network's nodes. The fine tuning of synaptic weights, global threshold and interactive feedback can maintain a high degree of complexity within stable domains of the parameters' space [22,23].

In this study, we first show that the attractor-based complexity of a Boolean network is related to the *discriminability degree* of specific input streams. We present a supervised attractor-based learning procedure aimed at achieving a maximal *discriminability degree* of a selected input stream. This procedure updates the *interactive weights* – the feedback projections which are combined with the external input – according to a target value of attractor-based complexity. We illustrate this learning procedure on a Boolean model of the basal ganglia-thalamocortical circuit. This circuit is known to be crucially involved in the processing and coding of information in the brain [24,25]. We discuss the possibility that the role of interactivity played by the modifiable feedback projections might correspond to the functional connectivity of the limbic system [26].

2 Attractor-Based Complexity of Boolean Recurrent Neural Networks

We briefly summarize the theoretical background exposed in detail in [18]. Any recurrent neural network $\mathcal{N}$ composed of Boolean integrate-and-fire (IF) units can be simulated by a corresponding finite state automaton $\mathcal{A}(\mathcal{N})$, and vice versa. The nodes of $\mathcal{A}(\mathcal{N})$ correspond to the different states (i.e., the spiking

configurations) of $\mathcal{N}$. There exists a transition from node i to node j labelled by u in $\mathcal{A}(\mathcal{N})$ if and only if $\mathcal{N}$ switches from state i to state j when receiving input u. Accordingly, the possible *dynamics* of a given network $\mathcal{N}$ correspond to the different *paths* in the graph of the associated automaton $\mathcal{A}(\mathcal{N})$.

Let us consider the Boolean network $\mathcal{N}$ of the basal ganglia-thalamocortical circuit (Fig. 1A) and its connectivity matrix [18] together with its corresponding finite automaton $\mathcal{A}(\mathcal{N})$ (Fig. 1B). Each node of the automaton represents a specific state of the network. For instance, node 384 corresponds to firing activity exclusively in the units representing the thalamus and superior colliculus (SC). If input $u = 0$ is received, which corresponds to unit 'IN' being not active, then the automaton switches from node 384 to node 223, which corresponds to activity in the units representing the thalamus, the thalamic reticular nucleus (NRT), the

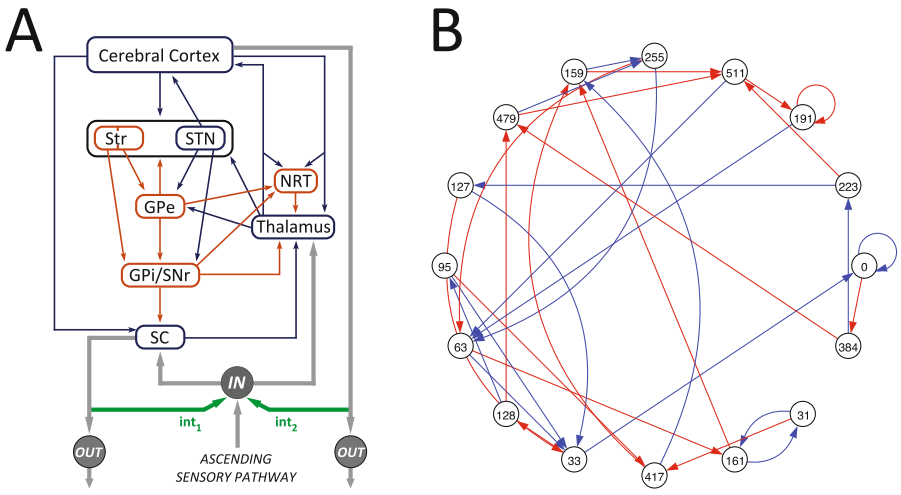

Fig. 1. A. Simplified Boolean model of the basal ganglia-thalamocortical circuit. Each brain area is represented by a single Boolean unit: superior colliculus (SC), Thalamus, thalamic reticular nucleus (NRT), Cerebral Cortex, the striatopallidal and the striatonigral components of the striatum (Str), the subthalamic nucleus (STN), the external part of the pallidum (GPe), and the output nuclei of the basal ganglia formed by the GABAergic projection neurons of the intermediate part of the pallidum and of the substantia nigra pars reticulata (GPi/SNR). We consider also the inputs (IN) from the ascending sensory pathway and the efferent outputs (OUT). The excitatory pathways are labeled in blue and the inhibitory ones in orange. We considered a closed-loop model with a recurrent connection from the efferent output to the input unit via 'interactive' connections int_1 and int_2. **B.** Finite automaton associated to the Boolean model of the basal ganglia-thalamocortical circuit. Each node of the automaton is a state of the circuit. There is a blue or red transition from node i to node j if and only if the network switches from state i to state j when receiving input 0 or 1, respectively. The cycles in the automaton correspond to the attractors of the network. (Color figure online)

pallidum (both internal GPi and external GPe parts), the striatum (Str), the subthalamic nucleus (STN) and the cerebral cortex. If the next input is $u = 1$, which corresponds to unit 'IN' being active, then the automaton switches from node 223 switches to node 511, which corresponds to firing activity in all units of the circuit. And so on, with each distinct node corresponding to a distinct unique activity pattern in the circuit.

According to this construction, the *attractors* of $\mathcal{N}$ – i.e., those dynamics which eventually become trapped into the repetition of a same set of states – correspond to the *cycles* in the graph of $\mathcal{A}(\mathcal{N})$. The set of attractors of network $\mathcal{N}$ can therefore be computed effectively, by constructing the associated automaton $\mathcal{A}(\mathcal{N})$ and listing the cycles of this latter. Note that any *periodic attractor* of $\mathcal{N}$ will necessarily elicit some recurrent *spatiotemporal pattern of discharges* which corresponds to the set of states visited periodically.

In this context, we introduced an *attractor-based measure of complexity* [18], which corresponds to the translation of a refined automata-theoretic notion [27] to the Boolean neural network context. Formally, suppose that $\mathcal{N}$ is a Boolean network provided with a classification of all of its attractors according to their meaningfulness – i.e., an attractor is classified as meaningful or spurious depending on the meaningfulness of its composing states (see [18] for more details). The *attractor-based complexity* of $\mathcal{N}$ is the integer n associated to a maximal sequence of cycles $\mathcal{C} = (C_0, \ldots, C_n)$ of $\mathcal{A}(\mathcal{N})$ – i.e., of attractors of $\mathcal{N}$ – which satisfies:

- C_i is included in C_{i+1}, for $i = 0, \ldots, n-1$; (1)
- C_i and C_{i+1} have opposite meaningfulness, for $i = 0, \ldots, n-1$. (2)

Conditions (1) and (2) state that the complexity measure is related to sequences of attractors that are included one into the next and of alternating meaningfulness. The general idea behind this complexity measure is that the meaningful and spurious attractors of a network are interpreted as the dynamical behaviors encoding the "acceptation" or "rejection" of the continual input received. Hence, a switch from one attractor to another of opposite meaningfulness corresponds to a moment when the network shifts from an "acceptation" to a "rejection" (or vice-versa) of its continual input. Accordingly, the attractor-based complexity of the network refers to its ability to discriminate between its input streams, by means of alternations between meaningful and spurious attractor dynamics [18]. This feature has been argued to be related to the computational power of the network (cf. [18] and Sect. 4).

The Boolean network of Fig. 1A has an attractor-based complexity of 6 with its connectivity matrix described elsewhere [18]. This value is highly dependent from both local and global variations of the synaptic strengths [22,23]. In fact, small perturbations of the connectivity weights and firing threshold might lead to completely distinct associated automata (with completely different cycle structures), and therefore, to very different attractor-based complexities. Furthermore, the interactive feedback (Fig. 1A, weights int_1 and int_2) plays a key role in the regulation of the network's attractor-based complexity. The parameter space defined by the variations of int_1 and int_2 shows the existence of *stable*

domains characterized by same values of the network's complexity, as illustrated by the different colored areas of Fig. 2.

Note that short input streams would induce the network's dynamics to visit only few (or no) attractors. By contrast, longer input streams will necessarily bring the network's dynamics into multiple successive attractors. For any such long input stream s, let $C_s = (C_0, \ldots, C_n)$ be a corresponding sequence of attractors visited by the network receiving input s (note that C_s is not unique). We will say that s is *discriminated* by $\mathcal{C}_s$ whenever C_s satisfies the above conditions (1) and (2). Accordingly, the *discriminability degree* of s, denoted as $d^*(s)$, is the largest number of attractor alternations that can be found in a sequence C_s which discriminates s. In other words, s has a *discriminability degree* of k if s is discriminated by some sequence $\mathcal{C}_s = (C_0, \ldots, C_k)$, but by no larger sequence $\mathcal{C}'_s = (C_0, \ldots, C_l)$ with $l > k$. Notice that by definition, if some input stream s has a discriminability degree of k in $\mathcal{N}$, then the attractor-based complexity of

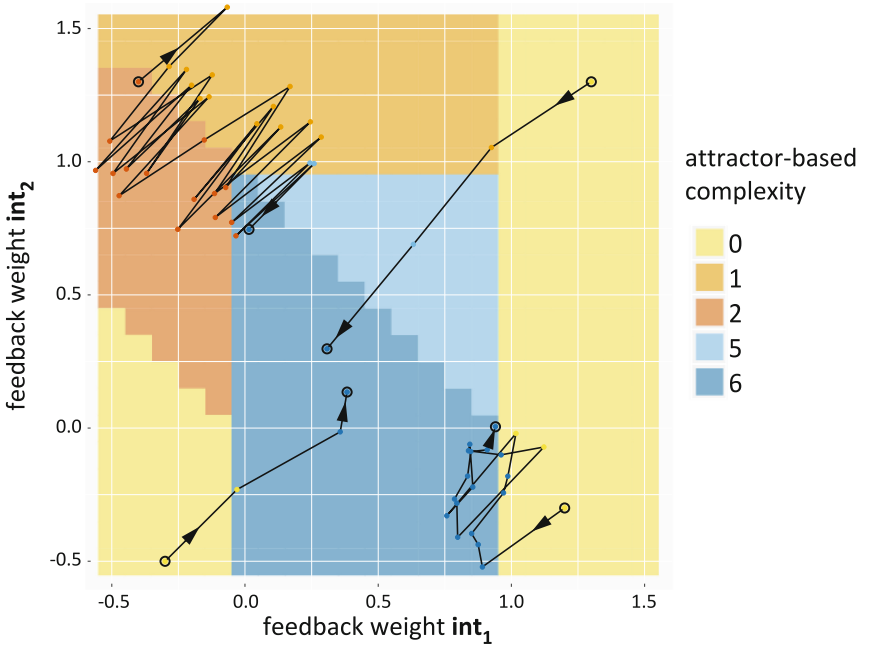

Fig. 2. Illustration of four trajectories of the attractor-based learning procedure. The color scale indicates the attractor-based complexity of the network of Fig. 1 as a function of its two interactive weights, with an optimal domain of complexity 6. Each trajectory describes a specific learning procedure updating the interactive weights at each step. The start and end points of the trajectories are the initial and final values of the interactive weights, and the intermediate points correspond to the successive updates of the weights achieved by the learning procedure. (Color figure online)

$\mathcal{N}$ is at least k (since $\mathcal{A}(\mathcal{N})$ contains at least the sequence $C_s = (C_0, \ldots, C_{k-1})$ discriminating s).

In the sequel, we will consider a specific input stream $\bar{s}$ having discriminability degree 6 for the circuit of Fig. 1A. Due to limited space available, we do not provide here the full description of that input.

3 Attractor-Based Learning

We consider a learning task consisting in the discrimination of a selected perceptual input fed into the basal ganglia-thalamocortical circuit via the ascending sensory pathway. The optimal learning is achieved if the dynamics associated with the reading of that perceptual input reaches the largest discriminability degree (in the sense of Sect. 2). In summary, the *attractor-based learning* procedure defined here performs updating of the network's weights with the aim of achieving a maximal discriminability degree of a selected input stream s. In our case, we assume that the connectivity matrix of the circuit is fixed and that the learning procedure only modifies the feedback weights int_1 and int_2 (Fig. 1A). The updating of the weights int_1 and int_2 depends on whether the network's dynamics induced by the given input stream s visits mainly spurious or meaningful attractors, and on the number of alternations between such attractors. The learning procedure is supervised in the sense that a target value for the discriminability degree is set at the begin.

More precisely, let s be an input stream, let w_k for $k = 1, \ldots, N$ be the modifiable weights of the network, and let N^* be the target value of the discriminability degree of s. Let C_s be a sequence of attractors visited by the network reading input s, and such that C_s contains a maximal subsequence that discriminates s. Let also $m_s \in \{-1, 1\}^{len(C_s)}$ be the "meaningfulness of C_s", simply defined as: $m_s(i) = -1$ if $C_s(i)$ is spurious and $m_s(i) = 1$ if $C_s(i)$ is meaningful. Finally, let $d^*(s)$ be the current discriminability degree of s. If the discriminability degree $d^*(s) < N^*$, the weights w_k are updated according to the following rule:

$$f(w_k) = w_k + step \cdot \frac{-sum(m_s)}{len(m_s)} \cdot \left(1 + \frac{len(m_s) - d^*(s)}{len(m_s)}\right) + \epsilon$$

where $sum(m_s)$, $len(m_s)$ are the sum and length of m_s, ϵ is a uniform noise in the range $[-0.1, 0.1]$, and $step = 0.3$. Note that if the reading of s induces a sequence C_s of only spurious (resp. of only meaningful) attractors, then $d^*(s) = 0$ and $|sum(m_s)| = len(m_s)$, and thus an update of maximal amplitude $f(w) = w + 2 \cdot step + \epsilon$ (resp. $f(w) = w - 2 \cdot step + \epsilon$) ensues. In words, the weight's update is increased when the number of alternations and the discriminability degree are lower (i.e., $|sum(m_s)|$ is high and $d^*(s)$ is small). The learning procedure based on this updating rule is given in Algorithm 1.

We illustrate this learning procedure in the case of the neural circuit presented in Sect. 2. For this purpose, we have considered $\bar{s}$ as the selected input

Algorithm 1. Attractor-based learning procedure

Require: input stream s; initial weights w_1, w_2; target discriminability degree N^*
1: compute $d^*(s)$
2: **while** $d^*(s) < N^*$ **do**
3: $w_k \leftarrow f(w_k)$, for $k = 1, \ldots, N$ `weights' updating`
4: compute $d^*(s)$ for the network with updated weights w_k, for $k = 1, \ldots, N$
5: **end while**
6: **return** w_k, for $k = 1, \ldots, N$

stream (cf. Sect. 2) and set the target discriminability degree to $N^* = 6$. We analyzed the learning procedure over the parameter space defined by the interactive weights int_1 and int_2 in the range $[-0.5, 1.5]$ by steps of 0.1 (Fig. 2). For each point in this space, we simulated the procedure from this point and followed its trajectory until it stopped. In the majority of the simulations (427/441), the procedure converged to novel interactive weights such that $d^*(\bar{s}) = N^* = 6$. Notice that during the procedure, the update of the weights int_1 and int_2 tended to be on the same direction (both increased or both decreased), which favored trajectories with angles between 30° and 60°.

4 Discussion

We have introduced an attractor-based learning procedure which modifies the modifiable weights of a network in order to achieve the optimal discrimination of a selected input stream. In our simplified model of the basal ganglia-thalamocortical circuit, we showed that interactive weights can be updated to reach a high level of discriminability of a given input stream, and in turn, to drive the dynamics of the network to a basin of attractions with a high level of complexity. Hence, the higher the level of discriminability, the larger the sequence of attractors visited by the dynamics, and accordingly, the larger amount of spatiotemporal patterns, and the higher the storage capacity of dynamic memories. We suggest that this correlation between attractor-based complexity and storage capacity of dynamic memories also prevails in real brain networks. Experimental evidence for bump attractor dynamics underlying spatial working memory has been provided by data from oculomotor delayed response tasks in awake behaving monkeys [28]. This study shows that persistent activity reinforcement is associated with a continuous prefrontal representation of memorized space, which is in agreement with other experimental data showing the emergence of recurrent spatiotemporal firing patterns associated with persistent activity in the inferotemporal cortex of behaving monkeys [11]. Hence, despite the oversimplification of our model (e.g. the brain probably is not behaving as a boolean network), the attractor-based complexity defined here may be considered an indicator of some aspects of the computational capabilities of neural networks.

General forms of synaptic plasticity and interactive architecture play a crucial role in regulating and controlling the computational and dynamical capabilities

of Boolean neural networks [22,23]. In the brain, the role assumed by the feedback might be played by the connections to and from the limbic system [26]. Such interaction reflects a dynamic adaptation to the learning situation. Dysfunctions of synaptic plasticity and functioning of the hippocampal formation and basal ganglia-thalamocortical loops may lead to impairment of learning, memory, and attention evoked by sleep deprivation.

References

1. Skarda, C.A., Freeman, W.J.: How brains make chaos in order to make sense of the world. Behav. Brain Sci. **10**, 161–173 (1987)
2. Tsuda, I.: Chaotic itinerancy as a dynamical basis of hermeneutics of brain and mind. World Futures **32**, 167–185 (1991)
3. Villa, A.E.P.: Empirical evidence about temporal structure in multi-unit recordings. In: Miller, R. (ed.) Time and the Brain. Conceptual Advances in Brain Research, vol. 3, pp. 1–61. CRC Press, London (2000)
4. Mpitsos, G.J., Burton, R.M., Creech, H.C., Soinila, S.O.: Evidence for chaos in spike trains of neurons that generate rhythmic motor patterns. Brain Res. Bull. **21**(3), 529–38 (1988)
5. Hoppensteadt, F.C.: Intermittent chaos, self-organization, and learning from synchronous synaptic activity in model neuron networks. Proc. Natl. Acad. Sci. U.S.A. **86**(9), 2991–2995 (1989)
6. Celletti, A., Villa, A.E.P.: Low-dimensional chaotic attractors in the rat brain. Biol. Cybern. **74**(5), 387–393 (1996)
7. Villa, A.E.P., Tetko, I.V., Celletti, A., Riehle, A.: Chaotic dynamics in the primate motor cortex depend on motor preparation in a reaction-time task. Cah. Psychol. Cogn. **17**, 763–780 (1998)
8. Segundo, J.P.: Nonlinear dynamics of point process systems and data. Int. J. Bifurcat. Chaos **13**(08), 2035–2116 (2003)
9. Abeles, M.: Local Cortical Circuits: An Electrophysiological Study. Studies of Brain Function, vol. 6. Springer, New York (1982)
10. Vaadia, E., Bergman, H., Abeles, M.: Neuronal activities related to higher brain functions-theoretical and experimental implications. IEEE Trans. Biomed. Eng. **36**(1), 25–35 (1989)
11. Villa, A., Fuster, J.: Temporal correlates of information processing during visual short-term memory. Neuroreport **3**(1), 113–116 (1992)
12. Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., Aertsen, A.: Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. Nature **373**(6514), 515–518 (1995)
13. Prut, Y., Vaadia, E., Bergman, H., Haalman, I., Slovin, H., Abeles, M.: Spatiotemporal structure of cortical activity: properties and behavioral relevance. J. Neurophysiol. **79**(6), 2857–2874 (1998)
14. Villa, A.E.P., Tetko, I.V., Hyland, B., Najem, A.: Spatiotemporal activity patterns of rat cortical neurons predict responses in a conditioned task. Proc. Natl. Acad. Sci. U.S.A. **96**(3), 1106–1111 (1999)
15. Asai, Y., Villa, A.E.: Reconstruction of underlying nonlinear deterministic dynamics embedded in noisy spike trains. J. Biol. Phys. **34**(3–4), 325–340 (2008)
16. Asai, Y., Villa, A.: Integration and transmission of distributed deterministic neural activity in feed-forward networks. Brain Res. **1434**, 17–33 (2012)

17. Iglesias, J., Villa, A.E.: Recurrent spatiotemporal firing patterns in large spiking neural networks with ontogenetic and epigenetic processes. J. Physiol. Paris **104**(3–4), 137–146 (2010)
18. Cabessa, J., Villa, A.E.P.: An attractor-based complexity measurement for boolean recurrent neural networks. PLoS ONE **9**(4), e94204 (2014)
19. Masulli, P., Villa, A.E.P.: The topology of the directed clique complex as a network invariant. Springerplus **5**, 388 (2016)
20. Cabessa, J., Villa, A.E.P.: The expressive power of analog recurrent neural networks on infinite input streams. Theor. Comput. Sci. **436**, 23–34 (2012)
21. Cabessa, J., Villa, A.E.P.: Expressive power of first-order recurrent neural networks determined by their attractor dynamics. J. Comput. Syst. Sci. **82**, 1232–1250 (2016)
22. Cabessa, J., Villa, A.E.P.: Attractor-based complexity of a boolean model of the basal ganglia-thalamocortical network. In: 2016 International Joint Conference on Neural Networks (IJCNN), pp. 4664–4671. IEEE, July 2016
23. Cabessa, J., Villa, A.E.P.: Attractor dynamics driven by interactivity in boolean recurrent neural networks. In: Villa, A.E.P., Masulli, P., Pons Rivero, A.J. (eds.) ICANN 2016. LNCS, vol. 9886, pp. 115–122. Springer, Cham (2016). doi:10.1007/978-3-319-44778-0_14
24. Nakahara, H., Amari Si, S., Hikosaka, O.: Self-organization in the basal ganglia with modulation of reinforcement signals. Neural Comput. **14**(4), 819–844 (2002)
25. Guthrie, M., Leblois, A., Garenne, A., Boraud, T.: Interaction between cognitive and motor cortico-basal ganglia loops during decision making: a computational study. J. Neurophysiol. **109**(12), 3025–3040 (2013)
26. Leblois, A., Boraud, T., Meissner, W., Bergman, H., Hansel, D.: Competition between feedback loops underlies normal and pathological dynamics in the basal ganglia. J. Neurosci. **26**(13), 3567–3583 (2006)
27. Wagner, K.: On ω-regular sets. Inf. Control **43**(2), 123–177 (1979)
28. Wimmer, K., Nykamp, D.Q., Constantinidis, C., Compte, A.: Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. Nat. Neurosci. **17**(3), 431–439 (2014)
29. Packard, M.G., Goodman, J.: Factors that influence the relative use of multiple memory systems. Hippocampus **23**(11), 1044–1052 (2013)
30. Lintas, A.: Discharge properties of neurons recorded in the parvalbumin-positive (pv1) nucleus of the rat lateral hypothalamus. Neurosci. Lett. **571**, 29–33 (2014)
31. Atallah, H.E., Frank, M.J., O'Reilly, R.C.: Hippocampus, cortex, and basal ganglia: insights from computational models of complementary learning systems. Neurobiol. Learn Mem. **82**(3), 253–267 (2004)
32. Perrig, S., Iglesias, J., Shaposhnyk, V., Chibirova, O., Dutoit, P., Cabessa, J., Espa-Cervena, K., Pelletier, L., Berger, F., Villa, A.E.P.: Functional interactions in hierarchically organized neural networks studied with spatiotemporal firing patterns and phase-coupling frequencies. Chin. J. Physiol. **53**(6), 382–395 (2010)