

A modular latching chain

Sanming Song · Hongxun Yao · Alessandro Treves

Received: date / Accepted: date

Abstract Many cognitive tasks involve transitions between distinct mental processes, which may range from discrete states to complex strategies. The ability of cortical networks to combine discrete jumps with continuous glides along ever changing trajectories, dubbed latching dynamics, may be essential for the emergence of the unique cognitive capacities of modern humans. Novel trajectories have to be followed in the multidimensional space of cortical activity for novel behaviours to be produced; yet, not everything changes: several lines of evidence point at recurring patterns in the sequence of activation of cortical areas in a variety of behaviours. To extend a mathematical model of latching dynamics beyond the simple unstructured auto-associative Potts network previously analysed, we introduce delayed structured connectivity and hetero-associative connection weights, and we explore their effects on the dynamics. A modular model in the small-world (SW) regime is considered, with modules arranged on a ring. The synaptic weights include a standard auto-associative component, stabilizing distinct patterns of activity, and a hetero-associative component, favoring transitions from one pattern, expressed in one module, to the next, in the next module. We then study, through simulations, how structural parameters, like those regulating rewiring probability, noise and feedback connections, determine sequential association dynamics.

Keywords Modular networks · Latching dynamics · Strategy change

Strategy transitions and latching

Transitions between discrete cortical states as well as between complex states unfolding in time, such as well-rehearsed segments (“routines”) of action plans, or even, at the opposite end, a stream of consciousness ([James 1892](#)), are thought to be an essential element of higher cognition. The ability to abandon the current state, or routine, or even schema, and jump elsewhere, can produce abstract thinking processes in an infinite variety of combinations. Take, for example, the so-called Stamma’s mate in chess. In Fig.1(a), when white moves the knight to b4 (*Nb4*), black has no choice but to move the king (*Ka1*). Following an automatic, well-rehearsed cliché, white might hunt the pawn by moving the king (*Kb3*), but then black could flee with its king (*Kb1*) and the game would end in a draw. By jumping, instead, to a novel strategy (the blue transition), white can move the king to c1 (*Kc1*), forcing black to advance with the pawn (*a2*), and then proceed to checkmate with the knight (*Nc2*).

Such internally generated strategy changes can be studied in the lab with the Hayling test ([Seyed-Allaei et al. 2010](#)). When subjects are required to complete a preconfigured sentence frame like “The ship sank very close to the ...” with an arbitrary word of their choice, but totally unrelated to the sentence and to its natural conclusion, usually they rapidly learn to shift to a new strategy that allows them to produce an unrelated response on most trials, such as the strategy of selecting an object around the lab. Supervised non-routine operations (*s*-operations) can thus be latched onto one

S. Song · H. Yao (✉)
School of Computer Science and Technology, Harbin Institute of Technology, 150001 Harbin, China
E-mail: sanmsoong@gmail.com, h.yao@hit.edu.cn

A. Treves
Cognitive Neuroscience, SISSA, via Bonomea 265, 34136 Trieste, Italy
E-mail: ale@sisssa.it

another, e.g. S1 to S2 in Fig.1(b), in the absence of additional external input, just like simpler routines. This departure from a standard procedure requires a particular type of “latching” (Treves 2005), latching between *s*-operations, which has been argued to be quintessential for the emergence of human intelligence (Amati and Shallice 2007). Another manifestation of latching processes may be seen in language generation, as suggested in Fig.1(d). Other examples, like the subtraction game, have been discussed by (Seyed-Allaei et al. 2010).

The latching Potts network considered by Treves (2005) is a simple unstructured autoassociative memory model with adaptive Potts units, that each represents a patch of cortex. Beyond its in-built functionality of storing and retrieving a large number of cortical activity patterns auto-associatively, it can also, in certain conditions, given adequate connectivity, latch from one pattern to the next in indefinitely long sequences (Russo and Treves 2012). Such sequences are neither random nor deterministic, rather they express an intermediate, complex matrix of transition probabilities between discrete states (Kropff and Treves 2005). Infinitely long sequences may potentially subserve the infinitely recursive processes at the core of the language faculty according to (Hauser et al. 2002). In its basic assumptions the Potts network, originally introduced by (Kanter 1988) as a purely statistical physics model, is compatible with widely held notions of cortical patches as functional units (Mountcastle 1998) with discrete activity states (Lansner et al. 2003) and sparse long range connectivity (Braitenberg 1978), patches which are themselves sparsely activated (Amit 1989). Discrete Potts states may stand for different color values (Kanter 1988), or edge direction, or texture type. Potts network models may also offer a bridge between macrolevel symbolic computation and microlevel neuronal dynamics, as increasingly invoked, e.g. in the domain of sentence parsing (Gerth and beim Graben 1998).

As it expresses transitions between simple discrete states with no intrinsic temporal dimension, the unstructured network, however, cannot seem to serve as a satisfactory model of transitions between routines, and even less between *s*-operations, both of which contain an internal dynamics that should be considered in the model. Moreover its connectivity, described in purely statistical terms (Fulvi-Mari and Treves 1998) does not produce sequences of activity patterns that can be related to observed patterns of cortical activation, thus forfeiting a potentially fruitful dialogue with experimental evidence. For example, episodic memory retrieval and prospective memory, critical elements in *s*-operations (Amati and Shallice 2007), appear to require the activation of anterior dorsolateral PFC regions and

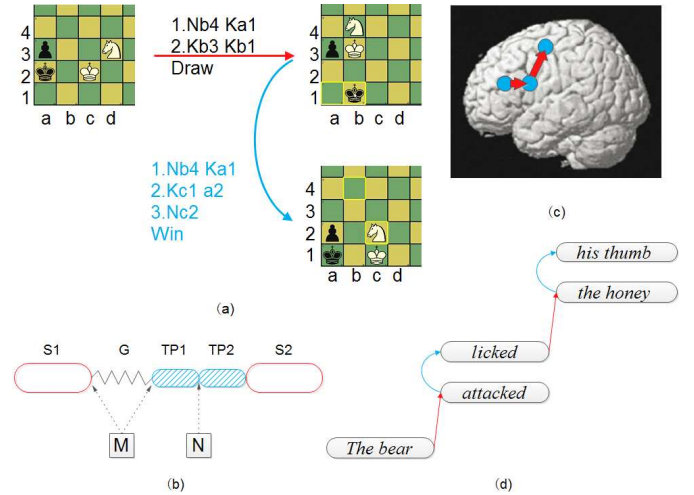


Fig. 1: Examples for strategy change. (a) A move in chess. White has to move. If the default strategy leads to a draw, a change in strategy is advisable. (b) Amati and Shallice (2007) illustrate the processes involved in performing the Hayling test in terms of (complex) latching, from the execution of the error-prone and effortful *s*-operation S1 to that of the successfully latched S2. *G* refers to the generation of a temporary goal by the *s*-operation, and *TP* is a transitional procedure. (c) Koechlin and Summerfield (2007) review the hierarchical patterns of activation involved in action and thought selection (or executive control), as seen with fMRI. Experimental results indicate that an ordered control signal arises along the anterior-posterior axis of the LPC (lateral prefrontal cortex). (d) A possible simplistic scheme for spontaneous sentence production. The subject (S) leads to a default verb (V1) by root association. A latching transition “localized” between verbs brings up another verb (V2, for example) in the absence of additional perceptual input. Similarly, further latching at the next stage picks a non-default object (O).

frontopolar cortex respectively. Imaging evidence (see Fig.1(c)) illustrates cases of hierarchically organized sequential activation patterns, which can be described by branching trees, with constant flow along the hierarchy and “decisions” (or perhaps transitions) at cortically localized branching points (Koechlin and Summerfield 2007).

Although the Potts network has an in-built small-world structure (Treves 2005), in that each unit represents a densely interconnected local network, there are indications from fMRI and MEG recordings (Bassett and Bullmore 2006) that cortical networks can be regarded as having a small-world architecture, also at the coarser level of connections through the white matter,

between distinct patches. A small-world connectivity would support stable hierarchical activation patterns, and it would affect network dynamics in a number of ways (Roxin et al. 2004; Guo and Li 2010), suggesting that it could be incorporated in a more structured extension of the original Potts model.

We present here an extended version of the model based on the definition of modules, at an intermediate level between the entire network and the individual units, which still represent patches of cortex. As a start, and to avoid issues with what happens at the top and bottom of cortical hierarchies, we consider five modules arranged on a ring, so that there is potentially information flow along the ring, with no beginning and no end. The connectivity between Potts units has a small-world structure, with denser connections within a module and sparser ones between modules. The connectivity between modules is made to encode a certain number of associations between pairs of patterns, modelling a root sequence deposited in long-term memory. Activity patterns tend to be expressed within one module at a time, although this is not enforced, but rather a product of the connectivity. Latching dynamics within each module can then be regarded as spontaneous transitions (including “strategy changes”) in each module, while hetero-associative connections (Li and Tsuda 2013) lead the network along the routine or *s*-procedure or hierarchy or branching tree or sentence, usually activating one module after the preceding one on the ring. A transmission delay is also introduced for intermodular connections, to eventually approach cortical plausibility.

We analyse such a modular latching chain by computer simulations. We focus on two measures, to characterize the basic traits of sequential association dynamics. One is the average latching chain length (LCL), i.e. the average number of sequentially recalled patterns in each module. LCL parametrizes the spontaneous productivity of the dynamics, that is, the number of transitions or alternatives considered before proceeding on with the next hetero-associative step. The other is the inverse switching ratio (ISR), i.e. the percentage of backward switching between sequential modules. ISR parametrizes the recourse to top-down or in general inverse information flow, interleaved with the standard feed-forward process. Here we report the main effects of structural parameters, like those regulating rewiring probability, noise and feedback connections, onto the sequential association dynamics described by these two measures.

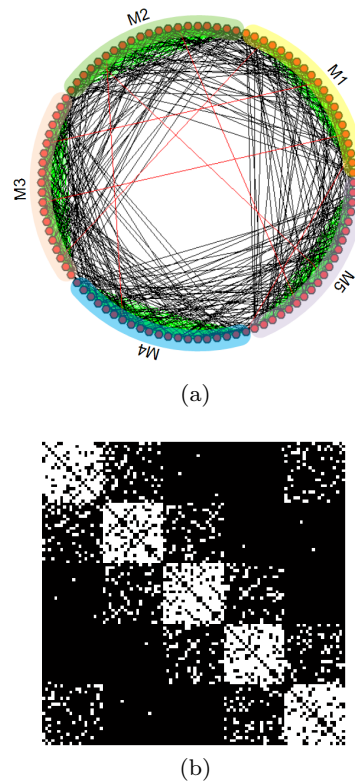


Fig. 2: A modular network of size $N = 100$ and $C = 20$ initial connections per unit. (a) The SW network after being rewired from a $C/2$ -nearest-neighbors regular network with rewiring probability $q = 0.01$ and uniform reshuffling within and between modules. Symmetrical auto-associative connections are in green and distant asymmetrical hetero-associative connections are in black (between neighbouring modules) or red (between distant ones). (b) The corresponding connection matrix, after adding feedback connections between modules, and making symmetric those within modules. Note that a feedback weight is weaker than its matching forward weight.

Model

It has been pointed out that a sparsely connected network of cortical patches has very limited storage capacity, unless two modifications are introduced (Fulviamari and Treves 1998). The first is non-uniform long-range connectivity. That is, connectivity is not uniformly sparse across patches, but rather it is concentrated between a patch and a subset of other patches that strongly interact with it. The second is sparse activity, at a macroscopic scale, which means that global activity patterns or semantic memory items are not defined across all patches, but only over a subset, different for each pattern, which tends to include strongly inter-

acting modules. The first of these two conditions for substantial memory capacity was not explicit in subsequent studies of Potts networks (Kropff and Treves 2005; Russo et al. 2008). The reduction of the *local* network in each patch to a single Potts unit, by itself, endows the *global* network with a large storage capacity. Here we stay with the Potts formulation, however connectivity is explicitly structured with a small-world (SW) scheme.

Modular Small-World architecture

To generate a modular SW network, we follow a standard (Watts and Strogatz 1998) procedure. It starts with a linear network with N Potts units arranged on a ring. Each unit is first connected to its nearest neighbors in a radius $C/2$ (with periodic boundary conditions), where C/N is the coefficient of “dilution” in the connectivity. Each edge or connection of the network is then randomly rewired with probability q . When $q = 0$, no rewiring occurs and the network remains a $C/2$ -nearest neighbors regular network; while for $q = 1$ it becomes a globally coupled (random) network, and the ring structure becomes irrelevant. For small q , the network displays SW characteristics. Both the average clustering coefficient and the mean shortest path length decrease gradually, as expected, with increasing q . With the simulation parameters in this paper ($N = 500$, and $C = 100$), the mean shortest path length varies by a small amount, due to the large connection radius $C/2$ (from 1.8637 for $q = 0$ to 1.7956 for $q = 1$, each with 100 simulations). Nevertheless, when $q = 0.1$, it is 1.7964, 1% larger than the minimum value, with respect to the range above, while the clustering coefficient (0.4000) is 67% above its minimum value (0.2074 for $q = 1$) with respect to the maximum (0.4949 for $q = 0$). One can therefore argue that for $q = 0.1$ (and for somewhat lower values) the network is in the small world regime.

After rewiring, the ring is partitioned into M modules, and connections within each module and between any given pair of modules are randomly reshuffled, to shed any dependence on the original position of the unit on the ring: only membership in a module remains relevant. The reason for this complex procedure is the need to later compare with a network with the same number of connections but a continuous, non modular ring structure, a comparison that will be presented elsewhere. An example of the final modular structure is illustrated in Fig.2(a).

The existence of a connection between a pair of Potts units is recorded in a binary matrix \mathcal{C} , with $c_{ij} = 1$ if there is a connection between units i and j , and

0 otherwise. The diagonal of the matrix is filled with zeroes. For example, Fig.2(b) shows the binary connection matrix of Fig.2(a).

Activity patterns and connection weights

In this paper we only consider the strictly modular network defined above, and in particular one with $M = 5$ modules. A different set of patterns are stored in each module, and we restrict our simulations to the case where $p = 10$ patterns are in each set, i.e. $pM = 50$ in all. Activity patterns are denoted in each module as ξ_i^μ , with $\mu = 1, \dots, p$ and $i = 1, \dots, N/M$, and they include a fraction a of the units assigned to any one of the S Potts active states, and the remaining fraction $(1 - a)$ assigned to the inactive state.

The network is constructed in the following way. Each module is defined as a standard auto-associative network (with Potts units), while across modules there are hetero-associative connections between pattern pairs, that facilitate transitions from one to the next, in a simple model of routines that have been deposited in long term memory. Note though that each pattern belongs to more than one routine. If pattern μ in module \mathcal{M} and pattern ν in module \mathcal{N} have been memorized as temporally consequent, they are called a “pattern pair” ($\mathcal{M}_\mu, \mathcal{N}_\nu$), or simply (μ, ν) . There are Ωp pairs between any two neighboring modules, that is, each pattern can lead through hetero-association to Ω other patterns in the next module. Given the periodic boundary conditions defining the ring (Fig.2), the “last” module is adjacent to the “first”.

We introduce these two kinds of synapses in the following two subsections. Auto-associative synapses are used to store patterns, in order to then enable completion of the correct pattern according to the input cues. And hetero-associative synapses are used to store associations between patterns in adjacent modules.

Auto-associative weights

Each Potts unit can be partially activated in a number S of distinct active states (a priori, with equal probability a/S) or else, in an inactive state (a priori, with probability $1 - a$). Activation levels in each state are graded, and at any time, their sum is normalized to 1, $\sum_{k=0}^S \sigma_i^k = 1$.

The Potts equivalent of the “Hebbian learning rule”, is written as a connection strength between unit i in

active state k and unit j in active state l ($i, j = 1, \dots, N$ while $k, l = 1, \dots, S$) (Kropff and Treves 2005)

$$\mathcal{J}_{ij}^{kl} = \frac{c_{ij}}{Ca(1-a/S)} (1 - \delta_{k0}) (1 - \delta_{l0}) \sum_{\mu=1}^p \left(\delta_{\xi_i^\mu k} - \frac{a}{S} \right) \left(\delta_{\xi_j^\mu l} - \frac{a}{S} \right) \quad (1)$$

where C is the number of connections arriving to Potts unit i (including those from other modules). δ is the Kronecker function, $\delta_{kl} = 1$ when $k = l$ and 0 otherwise.

Hetero-associative weights

Similar to the auto-association rule in Eq. (1), and to hetero-associative synapses in standard networks of binary units (Amit 1989), the hetero-association Hebbian learning rule when unit i in module \mathcal{M}_μ is in state k and unit j in module \mathcal{N}_ν is in state l can be written as

$$\mathcal{K}_{ij}^{kl} = \gamma \frac{c_{ij}}{2Ca(1-a/S)} (1 - \delta_{k0}) (1 - \delta_{l0}) \sum_{(\mu, \nu)} \left(\delta_{\xi_i^\mu k} - \frac{a}{S} \right) \left(\delta_{\xi_j^\nu l} - \frac{a}{S} \right) (1 - \delta_{\xi_i^\mu 0}) (1 - \delta_{\xi_j^\nu 0}). \quad (2)$$

where the last two factors denote the absence of contributions from patterns in the inactive state in either the pre- or post-synaptic unit. The factor γ regulates the strength of hetero- versus auto-associative weights.

Feedback connections of relative strength η are included, in order to assess their effect on network dynamics, chosen proportional to the forward connections,

$$\mathcal{K}_{ji}^{lk} = \eta \cdot \mathcal{K}_{ij}^{kl}. \quad (3)$$

and $0 \leq \eta \leq 1$.

Another element that we aim to assess is the presence of associations between non-neighbouring modules, which we call “noise pattern pairs”. The degree to which they are present, on average for each pattern, is denoted by ϵ , that is, a total $\epsilon p M$ pattern pairs are encoded in the weights between distant modules.

Uncorrelated patterns

A procedure to introduce model correlations in a group of p patterns is through a hierarchical algorithm, which may be parametrically varied from producing independent to highly correlated patterns (Treves 2005). In such a procedure, patterns are defined using a set of parents, from which they descend, emulating a genetic tree. These parents are defined simply as distinct random subsets of the entire set of N Potts units. In our simulations, however, for the sake of simplicity we have used only uncorrelated patterns. Each pattern is defined by randomly selecting a fraction a (see Formula (1)) of the units in a module, and randomly assigning one of the S active state to the unit in the pattern. The

remaining fraction $(1 - a)$ of the units are left in the inactive state. According to the estimation in Kropff and Treves (2005) (see their Fig.4), the theoretical storage capacity of each module for randomly correlated patterns is approximately $p_c \approx 0.035 C_m S^2 / a \approx 350$ for $C_m \approx 70$, $S = 6$ and $a = 0.25$, i.e. the parameters used in the simulations (C_m would be the final number of auto-associative connections within a module), so for any $p < 100$ each module should be well within its capacity. Due to the limited size of each module, however, even correlations between random patterns are very high, because of finite size fluctuations, so the actual capacity is far smaller than the theoretic bound. Considering also the limited computation resource, in the simulations we use only $p = 10$ patterns in each module.

Modular Latching Chain

Latching dynamics emerges as a consequence of incorporating two crucial elements in the Potts model: neuronal adaptation and correlation among attractors. Intuitively, latching may follow from the fact that all neurons active in the successful retrieval of some concept tend to adapt, leading to a drop in their activity and a consequent tendency of the corresponding Potts units to drift away from their local attractor state. At the same time, though, the residual activity of several Potts units can act as a cue for the retrieval of patterns correlated to the current global attractor. As usual with auto-associative memory networks, however, the retrieval of a given pattern competes, through an effective inhibition mechanism, with the retrieval of other patterns. In such a scenario, two conditions are fulfilled simultaneously, the global activity associated with a decaying pattern is weak enough to release in part the inhibition preventing convergence toward other attractors; but, as an effective cue, it is strong enough to trigger the retrieval of a new, sufficiently correlated pattern. In such a regime of operation, after the first, externally cued retrieval, the network concatenates in time successive memory patterns, i.e. it latches from attractor to attractor (Treves 2005).

In an auto-associative network without neural adaptation, the Potts states can be defined to be updated according to the heat bath rule,

$$\sigma_i^k(t+1) = \frac{\exp(\beta h_i^k(t))}{\sum_{l=0}^S \exp(\beta h_i^l(t))} \quad (4)$$

where $h_i^k(t)$ is a tensorial local “current” signal which sums the weighted inputs from other units, including

immediate auto-associative inputs and delayed hetero-associative inputs,

$$h_i^k(t) = \sum_{j \in \mathcal{M}_i} \sum_{l=0}^S \mathcal{J}_{ij}^{kl} \sigma_j^l(t) + \sum_{j \notin \mathcal{M}_i} \sum_{l=0}^S \mathcal{K}_{ij}^{kl} \sigma_j^l(t - \tau) + U \delta_{k0}(t) + w \left(\sigma_i^k(t) - \frac{1}{S} \sum_{l=1}^S \sigma_i^l(t) \right), \quad (5)$$

where in turn U is a threshold favoring the null state and τ is time delay between modules. The last term is a self-reinforcement term with coefficient w , which facilitates the convergence towards the more active state.

In our adapting network, instead, first we choose a very short numerical integration time, and then we set

$$\sigma_i^k(t) = \frac{\exp(\beta r_i^k(t))}{\sum_{l=0}^S \exp(\beta r_i^l(t))} \quad (6)$$

which is now mediated, for $k \neq 0$, by the vectors $r_i^k(t)$ (the “fields” which integrate the $h_i^k(t)$ “currents”) and by $\theta_i^k(t)$, the dynamic thresholds specific to each state, which are integrated in time

$$r_i^k(t+1) = r_i^k(t) + b_1 [h_i^k(t) - r_i^k(t) - \theta_i^k(t)] \quad (7)$$

and

$$\theta_i^k(t+1) = \theta_i^k(t) + b_2 [\sigma_i^k(t) - \theta_i^k(t)]. \quad (8)$$

While θ^k averages to σ^k in a typical time of b_2^{-1} steps, r^k averages to $h^k - \theta^k$ in a typical time of b_1^{-1} steps. We also include an overall threshold, i.e. effectively a non zero local field for the null state, driven by the integration of the total activity of unit i in all active directions,

$$r_i^0(t+1) = \theta_i^0(t) + b_3 [U + 1 - \sigma_i^0(t) - r_i^0(t)]. \quad (9)$$

Together with the threshold U , this local field for the null state regulates the unit activity in time, preventing local “overheating”. A threshold U of order 1 is crucial to ensure a large storage capacity (as shown by Tsodyks and Feigelman (1998)) and to enable unambiguous memory retrieval.

Results

All numerical simulations are performed in the MATLAB environment. In each run, a randomly chosen pattern is distorted by 20% noise (20% of the Potts units are set in a random state) and fed to the network as a partial cue. The initial “current” signal is used to initialize the local potential. The initial thresholds are set to zero.

When examining the effects of a parameter, we vary its value while keeping all others constant. The characteristics of the latching chains observed are measured

here by the average latching chain length (LCL) in each module (averaged across both modules and runs) and by the inverse switching ratio (ISR) of activity between modules. All values are calculated by averaging results of 20 successful independent simulations.

The general parameters setting is: network size $N = 500$, number of Potts states $S = 6$, number of modules $M = 5$ and sparsity parameter $a = 0.25$. Number of initial connections per unit $C = 100$, number of patterns in each module $p = 10$, number of pattern paired to each pattern $\Omega = 3$ and relative hetero-associative strength $\gamma = 0.1666$. Other parameters are temperature $\beta = 10$, self-reinforcement term $w = 1.8$, time delay $\tau = 1000$ steps. The time constants are $b_1 = 0.01, b_2 = 0.002, b_3 = 0.02$, with b_3 much larger than in the “slowly adapting” condition of (Kropff and Treves 2007), and consistent instead with the “rapidly adapting regime” discussed in (Russo and Treves 2012), chosen here primarily to accelerate the simulations. Note that the network remains in the resting state before external inputs arrive, because spontaneous activity is not introduced here.

Latching along the ring

A typical latching chain is shown in Fig.3 (a) and the corresponding pattern transition graph is displayed in Fig.3 (b). As we can see, in the example most of the time activity is concentrated in a single pattern in one module, and after lingering within the module for some time, possibly with transitions to other patterns, activity propagates to the next module.

Although it occurs for all practical purposes in continuous time (the integration time step is 50 times shorter than the shortest time constant), the propagation can be informally divided into several periods (marked by the dotted arrows). Module $\mathcal{M}1$ receives a partial cue A' at $t = t_0$, before which the global network is in the resting state. Then pattern A is retrieved as in a standard Hopfield auto-associative network. During the period $t_0 < t \leq t_0 + \tau$, other modules are still in resting states because of the synaptic transmission delay τ . Once the wave reaches module $\mathcal{M}2$ at $t = t_0 + \tau$, in $\mathcal{M}1$ activity attenuates, while it grows in $\mathcal{M}2$. If $\mathcal{M}2$ becomes active at t_1 , we can call $[t_0 + \tau, t_1]$ the activity transition period. In $t_1 < t \leq t_1 + \tau$, Potts units in module $\mathcal{M}2$ receive both auto-associative inputs from the same module and delayed hetero-associative inputs from $\mathcal{M}1$. Competitive effects drive the network either to latch to another pattern, following a correlation among patterns in $\mathcal{M}2$, or to proceed to a pattern in $\mathcal{M}3$ by hetero-association, as illustrated in Fig.3 (a). If

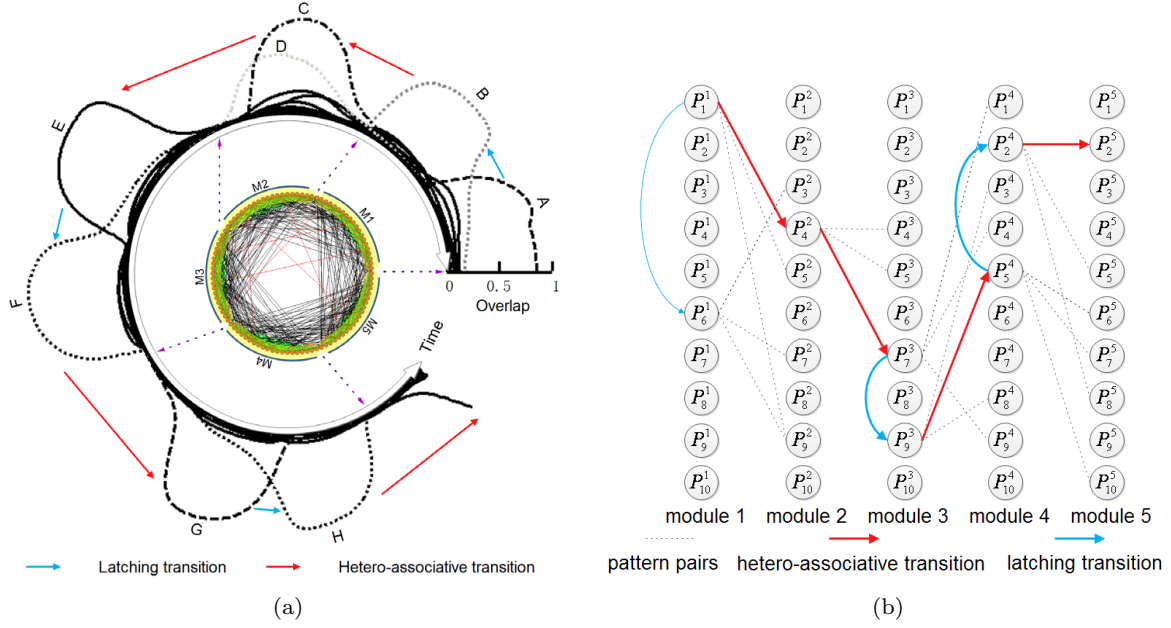


Fig. 3: (a) An example of a modular latching sequence. For easier visualization, the latching sequence is plotted along a circular time axis, with the network structure at the center. Purple dotted arrows mark the dynamical boundaries between modules. Red and blue arrows stand for hetero-associative and auto-associative (latching) transitions, respectively. (b) The corresponding pattern transition graph. Grey dotted lines are the stored pattern pairs.

latching occurs within a module, sometimes the hetero-association to the next module proceeds from the pattern latched to, as in $\mathcal{M}3$ and $\mathcal{M}4$ in Fig.3 (a); and sometimes (case $\mathcal{M}1$) from the earlier pattern, especially when latching occurs well into the transition period, the outcome of which is already largely determined.

In Fig.3(a), the number of retrieved patterns in each module is not the same. Occasionally, a latching transition is aborted, as the one to pattern D in $\mathcal{M}2$. There is constant competition between latching transitions and hetero-association. The result is mainly determined by the relative strength of pattern correlations. Latching transitions tend to occur when patterns exist strongly correlated with the current one. Otherwise, the next module is activated. Thus, correlations play the crucial role, within the limits set, essentially, by the delay parameter τ .

One should note that the number of patterns that may be activated by A is $\Omega = 3$, i.e. p_4^2 , p_5^2 and p_6^2 . Only pattern C (p_4^2) is activated in the end. The choice may be due to small fluctuations, effectively to noise.

No inverse transmission occurs in the case illustrated in the figure, so $\text{ISR} = 0$. As for LCL, in this sequence $\text{LCL} = (2 + 1 + 2 + 2)/4 - 1 = 0.75$ (one should subtract 1, to count only the number of genuine latching steps). With changes in the parameters, both LCL and ISR

can change widely. This is what we discuss in the last subsection of the Results.

Raising the threshold, and lowering it

One should appreciate that the relatively clean example illustrated in Fig.3 requires setting the appropriate parameters, and in particular setting the threshold U .

There is no latching when U is raised to too large a value. The pattern correlated with the input cue can be retrieved, if the cue is effective, but then the patterns itself is insufficient as a cue to overcome the threshold for the activation of another pattern, e.g. in the next module. In the example of Fig.4(a), with $U = 0.1$ and $q = 0.3$, it almost makes it, but not quite, and all activity stops on the ring after about 3000 steps.

An appropriate threshold value is illustrated in Fig.4(b), where $U = 0.075$ and $q = 0.3$. Then activity propagates along the ring, from module 2 to 1 to 5 to 4 to 3, raising one or a few patterns in each module.

If the threshold is lowered much further (and in order to show a clear example, we also slightly lower the rewiring probability, $U = 0.0375$ and $q = 0.2$), a new behaviour emerges, i.e. the activation of patterns in several modules at the same time. If the former can be denoted sL , or single-module latching, this would be mL

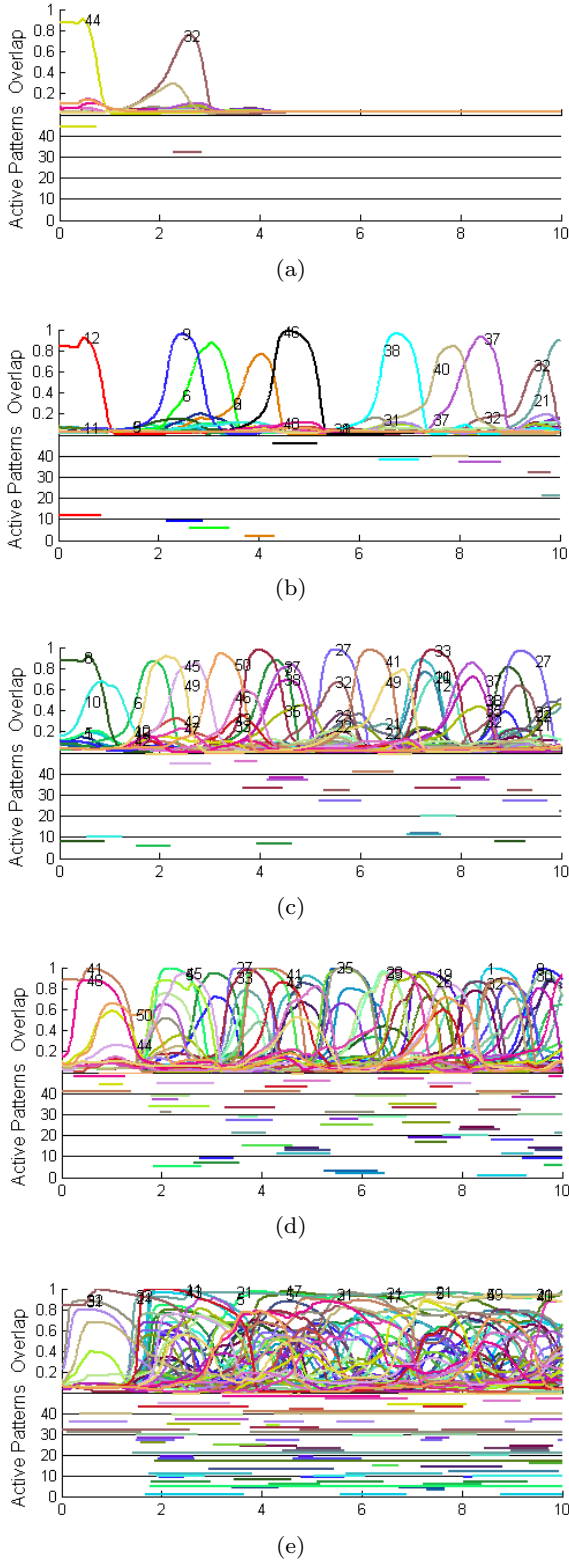


Fig. 4: Latching sequences from top to bottom, in the *NoL* (a), *sL* (b), *mL* (c), *infL* (d) and *SA* (e) phases, respectively. The location of these simulations in the phase diagram is displayed in Fig.5. Note that here $\eta = 0$ and $\epsilon = 0.5$. The unit for the time axis is 10^3 steps .

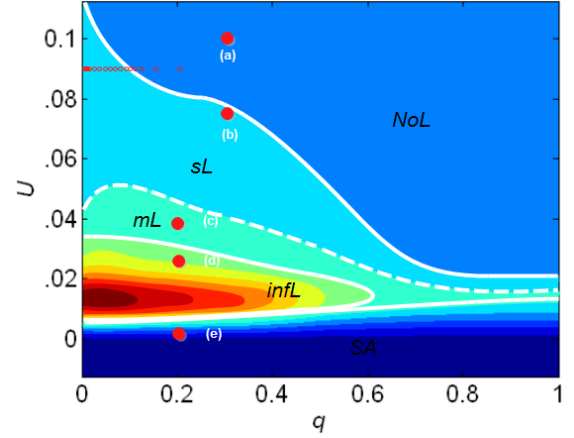


Fig. 5: The $U - q$ phase space, with approximate phase boundaries (solid curves) and, in false colors, contours of interpolated LCL values. In these simulations, $\epsilon = 0.5, \eta = 0$. *NoL*: No latching; *sL*: singular modular latching; *mL*: multi-modular latching; *infL*: infinite latching; *SA*: indefinitely stable attractors.

(for multi-module latching; Fig.4(c)). It is effectively a different phase for the system, in which the notion of propagation flow eventually loses meaning, and one can conceptualize it as a multiplicity of sequences intertwined in a sort of spaghetti.

When (U, q) are small, as in Fig.4(d), where $U = 0.025$ and $q = 0.2$, not only multiple modules are active at the same time, and multiple patterns are retrieved simultaneously in the same module, but also latching can carry on indefinitely, thus reproducing the infinite latching region (*infL*) of Russo and Treves (2012). With our parameters this occurs at lower U values than those yielding the (putative) transition from the *sL* to the *mL* phase, but whether this may be true in general is an open question.

As in the Russo and Treves (2012) analysis of a non-structured network, further facilitating retrieval, either by strengthening the local feedback w , as done there, or by lowering U further, as here, the system can get stuck in a steady attractor state (usually a mixture of many pure states). In the example of Fig.4(e), when $U = 0$ (and $q = 0.2$), the network shows simultaneously the alternation of some patterns and the permanence of others.

U - q phase diagram

The regions in parameter space where the different behaviours occur can be synthesized in a phase diagram. Two dimensional phase diagrams can be drawn by choosing different pairs of parameters, such as w and the

noise level T in Russo and Treves (2012). Here we choose the threshold U and the rewiring probability q , which is what parametrizes the connectivity range from a regular network to a random one. As a dependent variable we choose LCL, which largely covaries with the characteristics of the different phases of the system. Fig.5 shows that the different types of dynamics illustrated in Fig.4 occur in distinct regions of the phase diagram, which we draw, as for Fig. 4, after setting $\eta = 0$ (no feedback, for simplicity), and $\epsilon = 0.5$ (intermediate number of noise pairs).

To understand how network structure determines the dynamics, we set the threshold at a relatively high level, since for lower values structure becomes progressively irrelevant.

Rewiring probability

How does rewiring probability affect latching sequences? Since we see in Fig.5 that no latching occurs when q is large, we focus on its low values, by considering a logarithmic scale. We then assess, together with q , the effect of noise pairs, parametrized by ϵ and of feedback, parametrized by η . The average values of LCL and ISR for latching chains with different rewiring probability, and the four possible combinations of no feedback ($\eta = 0$) or moderate feedback ($\eta = 0.5$), and no noise ($\epsilon = 0$) or moderate noise ($\epsilon = 0.5$), are plotted in Fig.6. It appears from the plot that LCL values are relatively robust to noise and feedback, as long as either is present, whereas ISR is significantly enhanced, especially for quasi-regular networks ($q \rightarrow 0$) only when *both* are present. As a function of q , LCL is substantially enhanced at the center of the SW range, $10^{-2} \approx q \approx 10^{-1}$, while ISR is suppressed in the same range. The values of both measures reduce to zero for essentially random networks, $q > 0.2$.

Discussion

In conclusion, we have extended the notion of latching sequences to modular networks, with a prevailing direction of information flow determined by hetero-associative connectivity, and we have begun the study, through simulations, of how network parameters like rewiring probability, noise pattern pairs, feedback connections and thresholds affect latching dynamics. After lingering on a module for some time, with several latches, activity usually propagates to another module spontaneously. Auto-associative connections support free associative latching dynamics, while hetero-associative con-

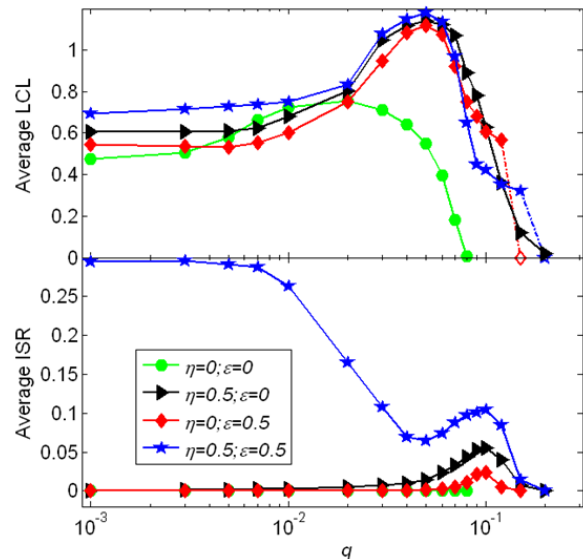


Fig. 6: The average LCL and ISR values as a function of rewiring probability q for four different levels of noise and feedback (and with $U = 0.09$). LCL is robust to the exact level of noise and feedback, as long as either is present. From the finite value, below one, that it takes for quasi regular networks, LCL is enhanced above one (provided noise *or* feedback are present) in the SW range, before reducing to zero when the network approaches random connectivity. Large ISR values instead appear to require both noise *and* feedback: when either is absent, there is only a limited region (closer to random than the SW range with nonzero probability of switching in the opposite direction), whereas when both are present the switching frequency has a trough in the SW range, before rising to a large value as the network approaches quasi-regular connectivity.

nections provide the pathways for attractor transitions between modules.

By varying the rewiring probability while keeping other system parameters fixed, the length of latching chains shows an inverted-U shape, indicating that a suitable rewiring probability, in the small-world range, enhances latching within modules. In approximately the same range, the probability of inverting the flow of activation, which is substantial only when both noise and feedback are present, has a relative minimum. Further, we find that both the LCL and ISR measures are to a certain degree insensitive, each in its own way, to the exact level of noise and feedback. We conclude that in a regime of high threshold, close to the maximum threshold for latching to occur, the network executes clean hetero-associative transitions along the ring, with

extensive latching transitions at each stage, when its connectivity is in the small-world regime.

To observe and analyze the neuronal network mechanism underlying strategy change, solely by experimental recordings (Seyed-Allaei et al. 2010), is highly non-trivial. Given the challenge, the use of computational models, while no replacement for experiments, may help guide thinking and stimulate new paradigms. An example close in spirit to our modular latching chain is the model put forward by Wennekers and Gunther (2009). Like our model, it uses hetero-associative connections to promote transitions between activity patterns. There are at least three important differences, though, between this work and their model. First, Wennekers' model is restricted to a single module while our model is comprised of several modules. Second, activity propagates solely following hetero-associative connections in their model, whereas auto-associative ones also exist, in ours, allowing it to execute both intra-class and inter-class pattern transitions. Third, the paradigm envisaged for Wennekers' model is to retrieve a fixed pattern sequence, albeit at a variable and manipulable pace, while our model generates an abundance of novel latching sequences at each run. Intra-modular latching dynamics, with their intrinsic productivity, generate sequences with no fixed order (Kropff and Treves 2007), while at the same time the "1-to-n" ordering of the hetero-connections provides a sketch of reproducible regional activation patterns in a variety of real-life tasks.

It is worth mentioning that Abeles (1982) has long ago proposed a "synfire-chain" model, which includes neuron pools similar to our modules. His model was originally intended to account mainly for the highly precise temporal patterns in neural firing times, that he had observed in macaque cortex. Constructed, again, with hetero-associative connections alone, it is likewise only able to retrieve a (stored) fixed pattern series, and it cannot be used to explain the strategy change phenomenon. In a sense, our modular latching chain integrates the merits of both "synfire-chains" and latching dynamics.

Sequential patterns of activity have indeed been widely observed in the mammalian cortex, and they have attracted considerable attention by scholars in recent years. The modular latching chains presented here are an attempt to simulate the generation mechanisms underlying complex cognitive processes. Further work on this topic is obviously needed, including at least the following three aspects. 1) Though we have begun to discuss the relationship between network structure and dynamics, choosing a parameter regime that closely relates to real cortical networks is an enduring endeavor, because dramatic changes may occur, vi-

a *phase transitions*, when parameters are only slightly tuned. 2) One goal of studying latching dynamics is to explore the mechanisms underlying infinite recursion and rule generalization in language (Pulvermüller and Knoblauch 2009). Russo et al. (2010) have pointed out that the dependence between successive patterns in a latching sequence approximates a second-order Markov chain, with a similar kinetics in a reduced artificial model of a natural language. So, on the one hand, we intend to study *generalization* in latching chains when a Hebb learning rule is adopted. On the other hand, it is important to analyse any analogy between latching chains and natural languages. 3) Strategy change appears to be a goal-guided behaviour. In this report, we have limited our analysis to the generation of free associative modular latching chains. How to control the latching process is waiting for future studies.

Acknowledgements S.S. thanks Eleonora Russo and Emilio Kropff for earnest and warmhearted assistance in sharing the idea of latching dynamics. S.S. and H.Y. are grateful for the support of the National Natural Science Foundation of China (61071180).

References

- Abeles, M (1982) Local Cortical Circuits: An Electrophysiological study. Springer, Berlin
- Amati D, Shallice T (2007) On the emergence of modern humans. *Cognition* 103:358–385
- Amit DJ (1989) Modeling brain function: The world of attractor neural networks. Cambridge University Press, New York
- Bassett DS, Bullmore E (2006) Small-world brain networks. *The Neuroscientist* 6:512–523
- Braitenberg V (1978) Cortical architectonics: general and areal. In P. H. Buzier MAB (Ed.) *Architectonics of the Cerebral Cortex*, pp. 443–465. Raven press, New York
- Fulvi-Mari CC, Treves A (1998) Modeling neocortical areas with a modular neural network. *Biosystems* 48:47–55
- Gerth S, beim Graben P (2009) Unifying syntactic theory and sentence processing difficulty through a connectionist minimalist parser. *Cognitive Neurodynamics* 3(4):297–316
- Guo D, Li C (2010) Self-sustained irregular activity in 2-d small-world networks of excitatory and inhibitory neurons. *IEEE Transactions on Neural Networks* 21:895–905
- Hauser M, Chomsky N, Fitch WT (2002) The language faculty: What is it, who has it, and how did it evolve? *Science* 298:1569–1579
- James W (1892) The stream of consciousness. *Psychology*
- Kanter I (1988) Potts-glass models of neural networks. *Phys. Rev. A* 37:2739–2742
- Koechlin E, Summerfield C (2007) An information theoretical approach to prefrontal executive function. *Trends Cogn. Sci.* 11:229–235
- Kropff E, Treves A (2005) The storage capacity of potts models for semantic memory retrieval. *Journal of Statistical Mechanics: Theory and Experiment* 8:P08010

- Kropff E, Treves A (2007) The complexity of latching transitions in large scale cortical networks. *Natural Computing* 2:169–185
- Lansner A, Fransén E, Sandberg A (2003) Cell assembly dynamics in detailed and abstract attractor models of cortical associative memory. *Theory in Bioscience* 122(1):19–36
- Li Y, Tsuda I (2013) Novelty-induced memory transmission between two nonequilibrium neural networks. *Cognitive Neurodynamics* 7(3):225–236
- Mountcastle VB (1998) *Perceptual Neuroscience: The Cerebral Cortex*. Harvard Univ. Press, Cambridge, MA
- Pulvermüller F, Knoblauch A (2009) Discrete combinatorial circuits emerging in neural networks: A mechanism for rules of grammar in the human brain? *Neural Networks* 22:161–172
- Roxin A, Riecke H, Solla SA (2004) Self-sustained activity in a small world network of excitable neurons. *Phys Rev Lett* 92(19):198101:1–4
- Russo E, Namboodiri VM, Treves A, Kropff E (2008) Free association transitions in models of cortical latching dynamics. *New J Phys* 10(015008):1–19
- Russo E, Pirmoradian S, Treves A (2010) Associative Latching Dynamics vs. Syntax. *Advances in Cognitive Neurodynamics(II)*:111–115
- Russo, E, Treves, A (2012) Cortical free association dynamics: distinct phases of a latching network. *Phys.Rev.E* 85(5):051920:1–18
- Seyed-Allaei S, Amati D, Shallice T(2010) Internally driven strategy change. *Thinking & Reasoning* 16(4):308–331
- Treves A (2005) Frontal latching networks: A possible neural basis for infinite recursion. *Cognitive Neuropsychology* 22(3-4):276–291
- Tsodyks M, Feigelman M (1998) The enhanced storage capacity in neural networks with low activity level. *Europhys Let* 6(2):101–105
- Watts D, Strogatz S (1998) Collective dynamics of 'small-world' networks. *Science* 286:509–512
- Wennekers T, Gunther P (2009) Syntactic sequencing in hebbian cell assemblies. *Cogn Neurodyn* 3:429–441