Neuronal mechanisms for sequential activation of memory items: dynamics and reliability

Elif Köksal-Ersöz[2,3], Carlos Aguilar[1,5], Pascal Chossat[2, 4], Martin Krupa[2,4], and Frédéric Lavigne[1]

[1]Université Côte d'Azur, CNRS-BCL, Nice, France
[2]Project Team MathNeuro, INRIA-CNRS-UNS, Sophia Antipolis, France
[3]Signal and Image Processing Laboratory, INSERM 1099, University of Rennes, Rennes, France
[4]Université Côte d'Azur, Laboratoire Jean-Alexandre Dieudonné, Nice, France
[5]Amaris, 950 Route des Colles, Biot, France

April 30, 2019

Abstract

In this article we present a biologically inspired model of activation of memory items in a sequence. Our model produces two types of sequences, corresponding to two different types of cerebral functions: activation of regular or irregular sequences. The switch between the two types of activation occurs through the modulation of biological parameters, without altering the connectivity matrix. Some of the parameters included in our model are neuronal gain, strength of inhibition, synaptic depression and noise. We investigate how these parameters enable the existence of sequences and influence the type of sequences observed. In particular we show that synaptic depression and noise drive the transitions from one memory item to the next and neuronal gain controls the switching between regular and irregular (random) activations.

1 Introduction

The processing of sequences of items in memory is a fundamental issue for the brain to generate sequences of stimuli necessary for goal-directed behavior [14], language processing [75] [13], musical performance [16] [20], thinking and decision making [21] and more generally prediction [17], [22], [23]. Those processes rely on priming mechanisms in which a triggering stimulus (e.g. a prime word) activates items in memory corresponding to stimuli not actually presented (e.g. target words; [12], [3]. A given triggering stimulus can generate two types of sequences: on the one hand, the systematic activation of a same sequence is required to repeat reliable behaviors [18], [24], [25], [26], [27]; on the other hand, the generation of variable sequences is necessary for the creation of new behaviors [28], [29], [30], [31], [33]. Hence the brain has to face two opposite constraints of generating repetitive sequences or of generating new sequences. Satisfying both constraints challenges the link between the types of sequence generated by the brain and the relevant biological parameters. Can a neural network with a fixed synaptic matrix switch behavior between reproducing a sequence and produce new sequences? And which neuronal mechanisms are sufficient for such switch in the type of sequence generated? The question addressed here is how changes in neuronal noise, short-term synaptic depression and neuronal gain make possible either repetitive or variable sequences.

Neural correlates of sequence processing involve cerebral cortical areas from V1 [27] [34] and V4 [26] to prefrontal, associative, and motor areas [35] [36]. The neuronal mechanisms involve a distributed coding of information about items across a pattern of activity of neurons [41] [42] [43] [44] [45]. In priming studies, neuronal activity recorded after presentation of a prime image shifts from neurons active for that image to neurons active for another image not presented, hence beginning a sequence of neuronal patterns [46] [47] [48] [49]. Those experiments report that a condition for the shift between

neuronal patterns of activity is that stimuli have been previously learned as being associated. Considering that the synaptic matrix codes the relation between items in memory [50] [51], computational models of priming have shown that the activation of sequences of two populations of neurons rely on the efficacy of the synapses between neurons from these two populations [53] [54] [55] [52] [12].

Turning to longer sequences, many of the models studied to date rely on the existence of steady patterns (equilibria) of saddle type, which allow for transitions from one memory item to the next [76] [8] [1]. Such models are well suited for reproducing systematically a same unidirectional sequence: as time evolves neuronal patterns are activated in a systematic order. These works show that the generation of directional sequences relies on the asymmetry of the relations between the populations of neurons that are activated successively. Regarding the order of populations n, n+1, n+2 in a sequence, the directionality of the sequence is obtained thanks to two properties of the synaptic matrix. First, the synaptic efficacy increases with the order of the populations, that is efficacy is weaker between populations one and two than between populations two and three [1] [18]. Second, the amount of overlap increases with the order of populations [8]. Indeed, individual neurons respond to several different stimuli [56] [57] [58] and two populations of neurons coding for two items can share some active neurons [59] [60]. Models have proposed a Hebbian learning mechanism that determines synaptic efficacy as a function of the overlap between the populations [61]. In models the amount of overlap codes for the association between the populations and determines their order of activation in a sequence [3] [4] [8] [1]. These works identify sufficient properties of the synaptic matrix to generate systematic sequences. However such properties of the synaptic matrix may not be necessary and neuronal mechanisms may also be sufficient to generate sequences.

In this work we consider the case of fixed synaptic efficacy and fixed overlap to focus on sufficient neuronal mechanisms that underlie the type of sequence. The present study mathematically analyses a new and more general type of sequences in which the states of the network do not need to reach saddle points. The model is based on a more general mechanism of transition from one memory item to the next, with the saddle pattern replaced by a saddle-sink pair (see Ashwin and Postelthwaite [2], for a prototype of this mechanism of transition). As time evolves the sink and saddle patterns become increasingly similar, so that even a small random perturbation can push the system past the saddle to the next memory item. In the model those new dynamics alleviate constraints on the synaptic matrix by allowing sequences that form spontaneously with the transitions obtained between populations related through fixed overlap, without theoretical or practical restriction on the length of the sequences. We investigate how changes in parameters with a clear biological meaning such as neuronal noise, short-term synaptic depression and neuronal gain can control the reliability of the sequences.

2 Results

2.1 The model

The focus of this paper is to present a mechanism of sequential activation of memory items in the absence of either increasing overlap, or increasing synaptic conductance, or any other feature forcing directionality of the sequences. We present this mechanism in the context of a simple system, however the idea is general and can be implemented in detailed models. We use the neural network model of the form

$$\dot{x_i} = x_i(1-x_i)\left(-\mu x_i - \lambda\sum_{j=1}^{N} x_j + \sum_{j=1}^{N} J_{i,j}^{max} s_j x_j\right) + \eta \tag{1}$$

$$\dot{s_i} = \frac{1-s_i}{\tau_r} - U x_i s_i \quad (i = 1, \ldots, N), \tag{2}$$

as in [1], with the variables $x_i \in [0,1]$ representing normalized averaged firing rates of excitatory neuronal populations (units), and $s_i \in [0,1]$ controlling short term synaptic depression (STD). The limiting firing rates $x_i = 0$ and $x_i = 1$ correspond respectively to the rest and excited states of unit i.

Any set $(x_1, \ldots, x_N)$ with $x_i = 0$ or 1 $(i = 1, \ldots, N)$ defines a steady, or equilibrium, *pattern* for the network. In the classical paradigm the learning process results in the formation of stable patterns of the network. Retrieving memory occurs when a cue puts the network in a state which belongs to the basin of attraction of the learned pattern.

Eq. (1) is usually formulated using the activity variable u_i (average membrane potential) rather than x_i, and x_i is related to u_i through a sigmoid transfer function. Our formulation in which the inverse of the sigmoid is replaced by a linear function with slope μ, was shown to be convenient for finding sequential retrievals of learned patterns, see [1].

The parameters in (1) are μ (or its inverse $\gamma = \mu^{-1}$ which is the gain, supposed identical, of the units, or slope of the activation function of the neuron [9]), λ the strength of a non-selective inhibition (inhibitory feedback due to excitation of interneurons) and $J_{i,j}^{max}$ the maximum weight of the connexion from unit j to unit i. Finally, η is a noise term which can be thought of as a flucutation of the firing rate due to random presence or suppression of spikes. In our simulations we considered white noise with the additional constraint of pointing towards the interior of the interval $[0, 1]$. Other types of noise can be chosen, this does not affect the mechanisms which we have invertigated.

Short-term depression reported in cortical synapses [77] rapidly decreases the efficacy of synapses that transmit the activity of the pre-synaptic neuron. This is modeled by eq. (2) where τ_r is the time constant of the synapse and U is the fraction of used synaptic resources. For an active unit with initially maximal synaptic strength $s_i = 1$, s_i decays towards the value $S = (1 + \rho)^{-1}$ where $\rho = \tau_r U$. The parameter ρ characterizes the STD.

The main difference in the model between this paper and [1] is the form of the matrix of excitatory connections $J^{\max}$:

$$J^{max} = \begin{bmatrix} 1 & 1 & 0 & \ldots & 0 \\ 1 & 2 & 1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 1 & 2 & 1 \\ 0 & \ldots & 0 & 1 & 1 \end{bmatrix}_{N \times N} . \tag{3}$$

This matrix is derived by the application of the simplified learning rule of [6] (details provided in [1]) using the collection of learned patterns

$$\xi^i = (0, \ldots, 0, 1, 1, 0, \ldots, 0) \ , \ i = 1, \ldots, P \tag{4}$$

where the two excited units are i and $i+1$. Conditions for the stability of these patterns in the absence of STD were derived in [1]. Note that the overlap between ξ^i and ξ^{i+1} is constant (one unit). By the application of the learning rule the coefficients of J^{max} are given by the formula:

$$J_{i,j}^{max} = \sum_{k=1}^{P} \xi_i^k \xi_j^k . \tag{5}$$

Consequently the matrix J^{max} is made up of identical (1 2 1) blocks along the diagonal, so that there is no increase in either overlap or the synaptic efficacy (weight) along any possible chain. We prove mathematically and verify by numerics that (1) admits a chain of latching dynamics passing through the patterns ξ^i, $i = 1, \ldots n-2$, either in forwards or in backwards direction depending on the activation, as well as shorter chains. The simplest way to switch dynamically from the learned pattern ξ^i to ξ^{i+1} is by having a mechanism such that unit i passes from excited to rest state, then unit $i+2$ passes from rest to excited state. STD can clearly result in the inhibition of unit i. However in the framework of [1] it was not possible to obtain the spontaneous excitation of unit $i+2$ with the connectivity matrix (3), because it was required that the upper and lower diagonal coefficients of J^{max} be strictly increasing with the order i. Connectionist models have shown the effects of fast synaptic depression on semantic memory [78] and on priming [3] [4]. Fast synaptic depression contributes to deactivation of neurons initially active in a pattern – because they activate less and less each other – in favor of the activation of neurons active in a different but overlapping pattern – because

newly activated neurons can strongly activate their associates in a new pattern. The combination of neuronal noise and fast synaptic depression enables latching dynamics in any direction depending on the initial bias due to random noise. Indeed, when the parameters lie within a suitable range, the action of STD has the effect of creating a "dynamic equilibrium" with a small basin of attraction. This dynamic equilibrium could be ξ^i, $\hat{\xi}^i$ (the pattern in which only unit $i+1$ is excited) or an intermediate pattern for which the value of x_i is between 0 and 1. Subsequently the noise allows the system to eventually jump to ξ^{i+1}, the process being repeated sequentially between all or part of the learned patterns. This noise-driven transition is what we call an *excitable connection* by reference to a similar phenomenon discussed in [2]. Chains of excitable connections can also be activated or terminated by noise. Last but not least we show that our system, depending on the neuronal gain γ, will follow the sequence indicated by the overlap or execute a random sequence of activations. Changes in neuronal gain change the sensitivity of a neuron to its incoming activation ([9], [10], [11]), and are reported to impact contextual processing [80] to enhance the quality of neuronal representations [79] and to modulate activation between populations of neurons to reproduce priming experiments ([40]). Here we show how changes in neuronal gain switches the networks behavior between repetitive (reliable) sequences and variable (new) sequences.

We proceed to present the results in more detail, as follows. In Section 2.2 we present simulations for the network with $N = 8$, which serves as an example of the more general construction. In Section 2.3 we sketch the methods we use to search for or verify the existence of the chains. In Section 2.4 we discuss irregular chains of random activations versus regular chains defined by the overlap.

2.2 Case study: a system with $N = 8$ excitable units

We consider sequences of seven learned patterns $\xi^1, ..., \xi^7$ (named ABCDEFG) encoded by eight units $x_1, ..., x_8$. The sequence represents the sequential activation of pairs of units 1-2, 2-3, 3-4, 4-5, 5-6, 6-7 and 7-8, corresponding to patterns A and B, B and C, etc. with an overlap of one unit between them. Learning is reported to rely on changes in the efficacy of the synapses between neurons [81] through long term potentiation (LTP) and long term depression (LTD) [82] [83] [84]. As a consequence, LTP/LTD potentiates/depresses synapses between units coding for patterns as a function of their overlap, that is synapses between units coding for overlapping patterns are more potentiated. Due the constant overlap, all synapses between overlapping patterns are equal. Note that the matrix J^{max} (3) is learned as a function of the overlap between patterns without imposing any sequences. A consequence is that learning of independent pairs of patterns generates a matrix that allows for the activation of sequences.

A system of $N = 8$ excitatory units can encode $P = N - 1 = 7$ regular patterns in J^{max} (3). Encoded memory items can be retrieved either spontaneously (noisy environment) or when the memory network is triggered by an external cue [8] [7]. Units x_1 and x_8 are the least self-excited units with $J^{max}_{1,1} = J^{max}_{8,8} = 1$, thus it is very unlikely to active them unless they are part of the initial activity state. Hence, the longest chain has $P - 1 = 6$ consecutive patterns.

2.2.1 Directional sequences from a stimulus-driven pattern in the sequence

Starting from the first pattern A, the directional activation corresponds to the sequence ABCDEFG (Fig. 1-a left panel). The forward direction is imposed by J^{max} because x_1 is less excited since $J^{max}_{1,1} = 1$. Hence, while the synaptic variables s_1 and s_2 are equal and decreasing together as the system lies in the vicinity of ξ^1, x_1 is deactivated before x_2. In the same interval of time $s_2 < s_3$ and $s_2 - s_3$ increases so that x_2 becomes unstable before x_3 and the system now may converge to ξ^2. The process can be repated between ξ^2 and ξ^3 etc. Similarly, starting from the last pattern G gives the reverse direction (GFEDCBA) to the system (Fig. 1-b right panel).

Initialising the system from a middle pattern ξ^i doesn't introduce any direction, since the two active units of ξ^i are equally excited. While their synaptic variables are decreasing together, depending on the noise at the moment when ξ^i becomes unstable, either ξ^{i-1} or ξ^{i+1} is activated with equal probabilities. Figure 2 shows the response of the system starting from a mid-point pattern D. The

activated sequence can go in either direction DEFG or its reverse DCBA. The *random* choice for a sequence is driven by a bias in the noise at the time of stimulus-driven activation of the mid-point pattern.

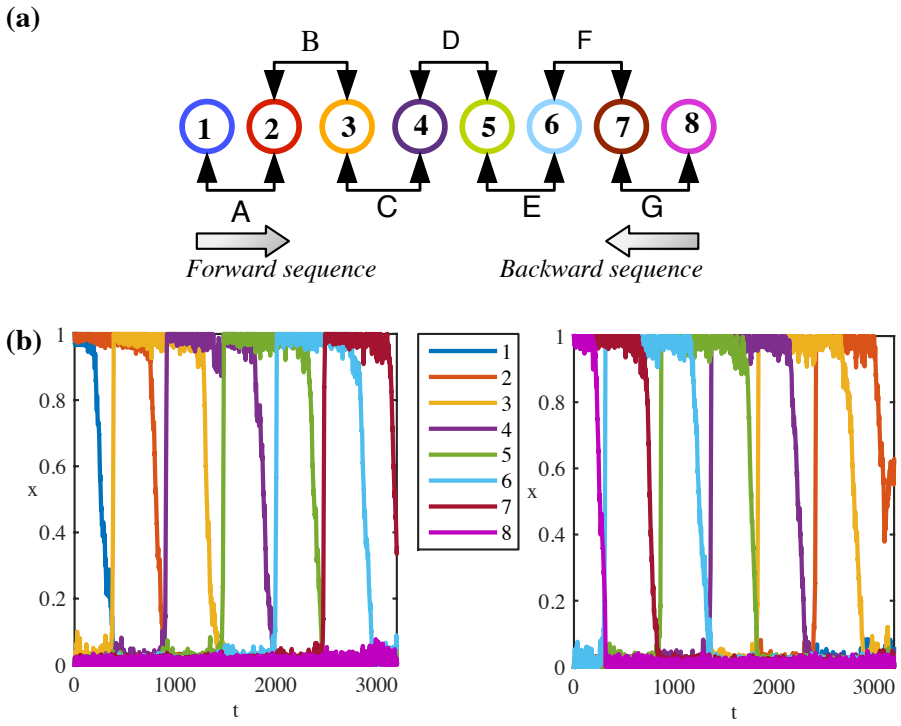

Figure 1: Directional sequences of an endpoint stimulus-driven system. **(a)** Each numbered circle represents a neuron. Consecutive units encode a pattern. Except for x_1 and x_8, each unit participates in two patterns. A forward sequence is the activation of units in increasing order. A backward sequence is the activation of units in decreasing order. **(b)** Left panel: System initialised from the pattern A follows the forward sequence until the pattern F. Right panel: System initialised from the pattern G follows the backward sequence until the pattern B. Same colour code is used to represent units' indices in **(a)** and **(b)**. Parameters: $\mu = 0.40714, \lambda = 0.50714, I = 0, \tau_r = 900, U = 0.002, \eta = 0.02$.

2.2.2 Noise-driven random sequence from a mid-point pattern in the sequence

The units that participate in two patterns (overlapping units) have stronger self-excitation as it is manifested by the diagonals of J^{max}. These units $(x_i, i \neq \{1, 8\})$ are likely to be excited by random noise and they can activate others that they encode a pattern with. After a pattern ξ^i or the associated intermediate pattern $\hat{\xi}^i = (0, \ldots, 0, 1, 0, \ldots, 0)$ being randomly excited by noise, the system can follow either ξ^{i-1} or ξ^{i+1}. The robustness of activity depends on the system parameters. Figure 3 shows an example of spontaneous activation of a mid-point pattern D where the directional oriented sequence can be either DEFG or its reverse DCBA. Similar to the system initialised from a middle pattern, the *random* choice for a direction is driven by a bias in the noise at the time of noise-driven activation.

2.2.3 Sensitivity of the dynamics upon parameter values

We have seen that patterns can be retrieved sequentially when the system is triggered by a cue or spontaneously by noise. However the effectiveness of this process depends on the values of the parameters in equations (1)-(2). The dynamics of the system can follow part of the sequence, then either terminate on one pattern ξ^i with $i < N - 1$, or converge to a non learned pattern. Moreover we identified two different dynamical scenarios by which a sequence can be followed, depending mainly on the value of μ (or gain $\gamma = \mu^{-1}$). This will be analyzed in Section 2.3. Here we comment on numerical simulations which highlight the dependency of the sequences upon parameter values.

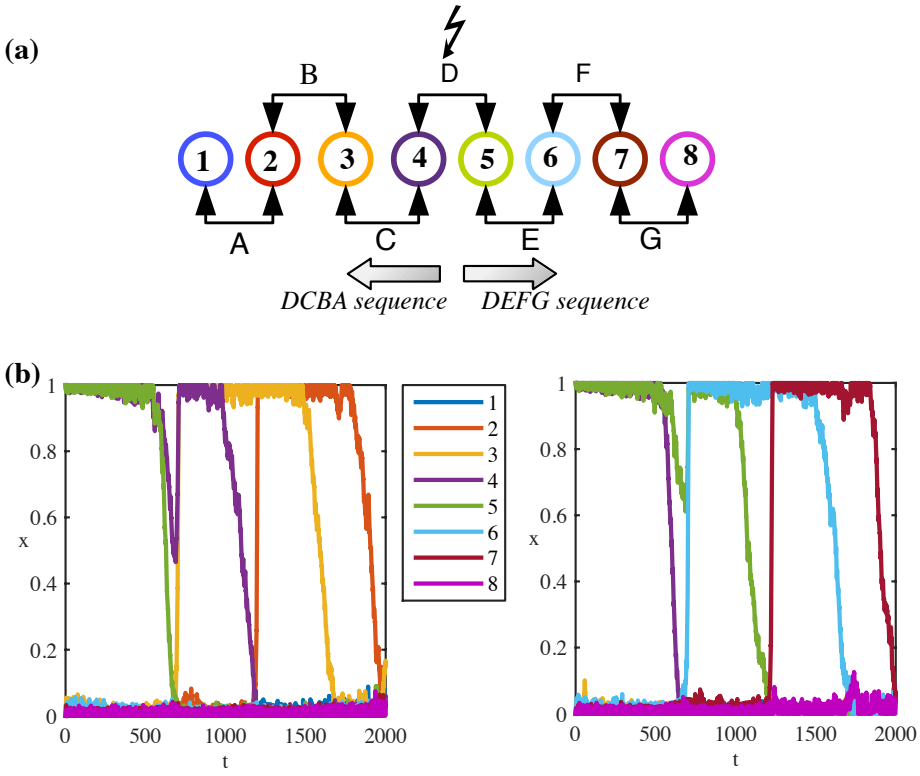

Figure 2: Directional sequences of a midpoint stimulus-driven system. **(a)** Each numbered circle represents a unit. Consecutive neurons encode a pattern. Except for x_1 and x_8, each unit participates in the encoding for two patterns. When the system initialised from the pattern D, it follows either the "DCBA" or "DEFG" sequence. **(b)** Left panel: System initialised from the pattern D follows the "DCBA" sequence until the pattern B. Right panel: System initialised from the pattern D follows "DEFG" sequence until the pattern F. The same colour code is used to represent units' indices in **(a)** and **(b)**. Parameters: $\mu = 0.40714, \lambda = 0.50714, I = 0, \tau_r = 900, U = 0.002, \eta = 0.02$.

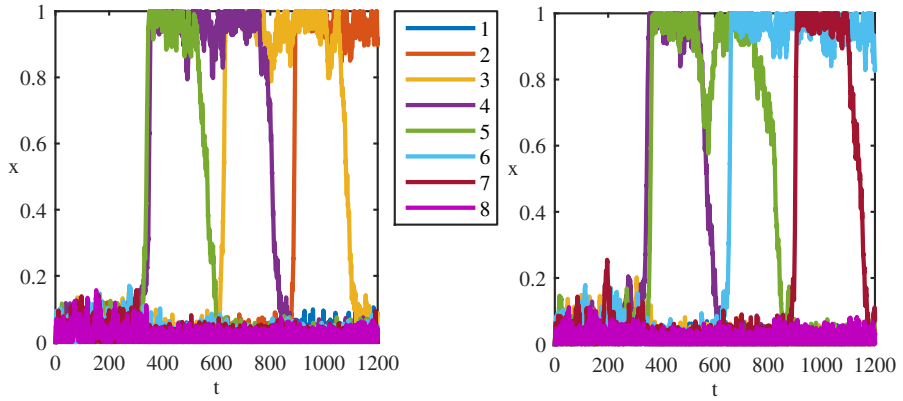

Figure 3: System activated spontaneous by random noise can move in backward (left panel) or forward (right panel) directions. Parameters: $\mu = 0.20714, \lambda = 0.50714, I = 0, \tau_r = 300, U = 0.006, \eta = 0.04$.

Figures 4abc show time series of the full or partial completion of sequences of retrievals (for $N = 8$ units) for two different values of noise amplitude $\eta = 0.02$ (first row) and $\eta = 0.04$ (second and third row). In each case the two first columns show statistics with STD parameters $\tau_r = 300$ and $U = 0.006$ while the two last columns correspond to the choice $\tau_r = 900$ and $U = 0.002$. Recall that τ_r gives the speed at which synaptic variables decay. By taking the product $\rho = \tau_r U$ constant we ensure that the synaptic variables s_i decay to the same value S in both cases (see 2.1). In 4a, 4b the global inhibition coefficient λ is set at 0.50741 in row two and $\lambda = 0.55741$ in row three. For each choice of STD parameters the value of μ differs between the left and the right columns: $\mu = 0.40714$ on the

left column and $\mu = 0.20714$ on the right column for Figs 4a, 4b, $\mu = 0.35714$ on the left column and $\mu = 0.15714$ on the right column for Fig 4c. Color indicates the activity of each unit from 1 to 8.
Observe that the sequence and the pattern durations are shorter in the system with fast synapses ($\tau_r = 300$) than the one with slow synapses ($\tau_r = 900$). In the case of a weaker noise (Figure 4a) with fast synapses the system follows the sequence ABCDEF when $\mu = 0.40714$ whereas it stops at the pattern B when $\mu = 0.20714$. In other words, increasing μ in the system with fast synapses tends to recruits neurones sequentially. The system with slow synapses can follow the sequence ABCDEF in both cases. In fact the two different values of μ in each row correspond to the two different dynamical scenarios which have been evoked in the beginning of this section. This point will be developed in Section 2.3. When noise is stronger (Figure 4b) the picture is different: the full sequence can be completed in the case of fast synapses even with $\mu = 0.20714$, however in the case of slow synapses the sequence is short and the system quickly explores unexpected patterns like one with three excited units $4, 5, 6$ around $t = 1000$ (which is not a learned pattern) in panel three.
Comparison between Figures 4b and 4c examplify the effect of changing λ and μ for the same noise amplitude. Increasing the inhibition coefficient λ regulates the transition for slow synapses, however fast synapses and high values of μ randomly activates learned patterns and yields short sequences (when however μ is smaller regular sequences can be preserved).

2.2.4 Length of a chain

When the patterns in a chain are explored in the right order by the system we call it *regular*. As we saw in Section 2.2.3 it can happen that only part of the full regular chain has been realised before it stops or starts exploring patterns in a different order, hence activating an irregular chain. We call the partial regular chain a *regular segment* and its length is the number of patterns it contains. Here we investigate the maximal length that a regular segment starting at pattern A can attain. This length is the rank of the last activated pattern over simulations. It depends on noise but also on the neuronal and synaptic parameters. In Figures 5 and 6 we present statistics of this rank for two different noise intensities $\eta = 0.02$ and 0.04. Last activated patterns are represented by color bars, the length of which indicates the percentage of corresponding last activated pattern over 80 trials. In each figure we took $\rho \in \{1.2, 2.4\}$,$\tau_r \in \{300, 900\}$. Parameters λ and μ are varied wihin a range assuring the existence of chains of at least length 2.
Generally speaking, increasing the noise level facilitates the activation, hence, prolongs the chains. Especially for $\tau_r = 900$ the chain length is considerably higher with $\eta = 0.04$. We can also observe the difference between $\rho = 1.2$ and $\rho = 2.4$, for the former strong inhibition is more favourable but week inhibition is more favorable for the later .
Also note that when μ is small, an increase of μ provokes an increase of the length of the chain. However, in most cases we find that the chain lengths are maximal for intermediate values of μ. This is clear intuitively: large gain (small μ) prevents the units from deactivating, making the transition from one pattern to the next difficult. Small gain, on the other hand, prevents the next unit from activating. Another factor is the occurrence of the transition from scenario 1 to 2 (see Section 2.3 and Appendix A.1). The noise level ρ and the global inhibition parameter λ also influence the system's behaviour. For $\rho = 1.2$, inhibition in the middle range leads to longer sequences, whereas weak inhibition is more suitable for $\rho = 2.4$.

2.3 Analysis of the dynamics

Latching dynamics is defined as a sequence (chain) of activations of learned patterns that de-activate due to a slow process (e.g., adaptation, here synaptic depression), allowing for a transition to the next learned pattern in the sequence ([5], [3]).
Here we refine this description using the language of dynamics and multiple timescale analysis. The main idea is to treat the synaptic variables s_i as *slowly varying parameters*, so that the evolution of the system becomes a *movie* of the dynamical configurations of the units x_i. On the other hand the firing rate equation (1) is well adapted to analyze latching dynamics. Indeed from the form of (1) (assuming for the moment that noise is set to 0) one can immediately see that whenever x_i is set to

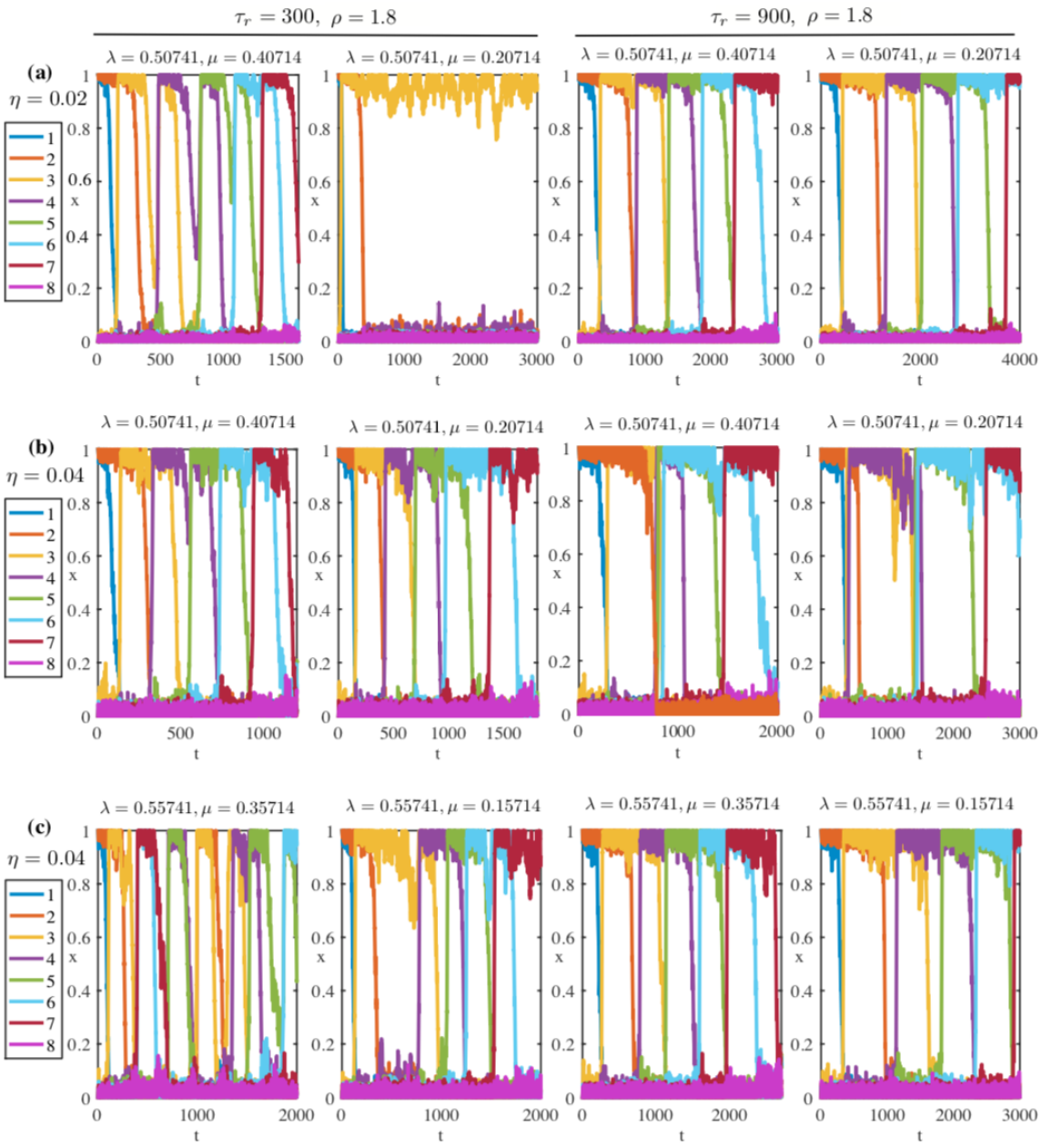

Figure 4: Response of the system initialised from pattern A to different levels of noise. Synaptic variables are faster along the first two columns ($\tau_r = 300$) than the last two columns ($\tau_r = 900$). **Row-(a)** The system with fast synapses and weak perturbation ($\eta = 0.02$) can follow the longest sequence from A to F for $\mu = 0.40714$ but not for $\mu = 0.20714$, while the slow synapses can trigger the longest sequence in either case. Increasing the noise amplitude to $\eta = 0.04$ (**row-(b)**) enables the activation of the whole sequence but slow synapses give very short patterns or 3 co-active units. While increasing the inhibition (**row-(c)**) regulates the transition for slow synapses, the system with high values of μ and fast synapses randomly activates learned patterns and yields short sequence. On the other hand, the system with small values of μ and fast synapses can preserve a regular sequence.

0 or 1, this variable stays fixed at any time. Therefore considering any face in the hypercube $[0,1]^N$ defined by two coordinates x_i, x_j, the other coordinates being fixed at 0 or 1, it is invariant under the flow of (1). In other words any trajectory starting in F stays entirely in it. This is of course true also for the edges and vertices at the boundary of each face. Each vertex is an equilibrium of (1) and connections between such equilibria can be realised through edges of the hypercube, which greatly simplifies the analysis.

When the couple (x_i, s_i) of unit i is set at $(1,1)$, x_i is fixed as we have seen but STD equation (2)

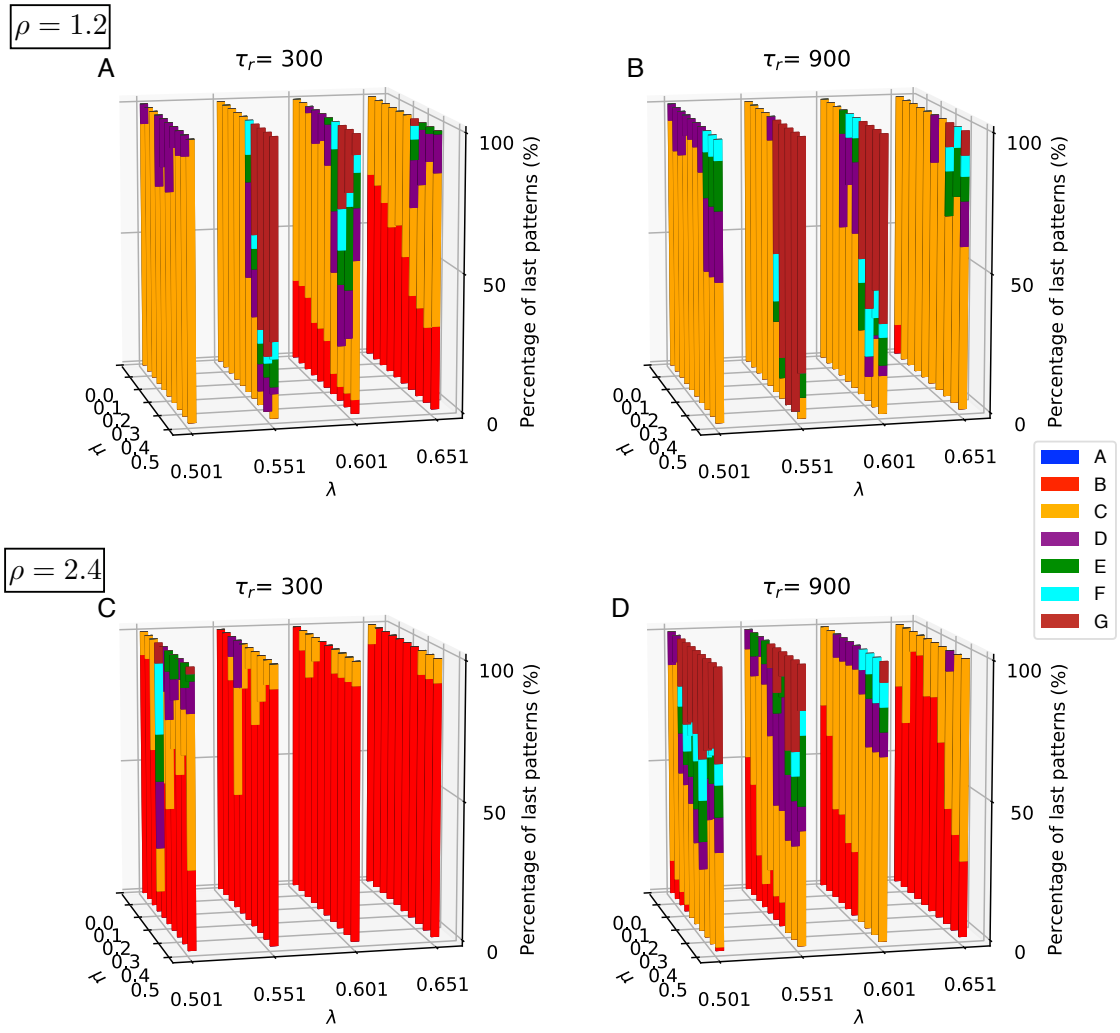

Figure 5: Percentage of last activated patterns in a regular segment over simulations for $\eta = 0.02$. Pattern colours follow to the colour codes of the last activated units (see the legend on the right). The height of each colour on a bar indicates the percentage of corresponding last activated pattern. For $\rho = 1.2$ (panels A and B), the chain length increases with (μ, τ_r). The global inhibition, λ, should be high enough for a sequential activation, but the chain length decreases if the inhibition is too strong. For $\rho = 2.4$ (panels C and D), the chain length increases with (μ, τ_r), but decreases with λ. The sharp increase in the chain length (more visibly in $\rho = 1.2$) occurs when the bifurcation scenario changes around $\mu \approx \mu^*$.

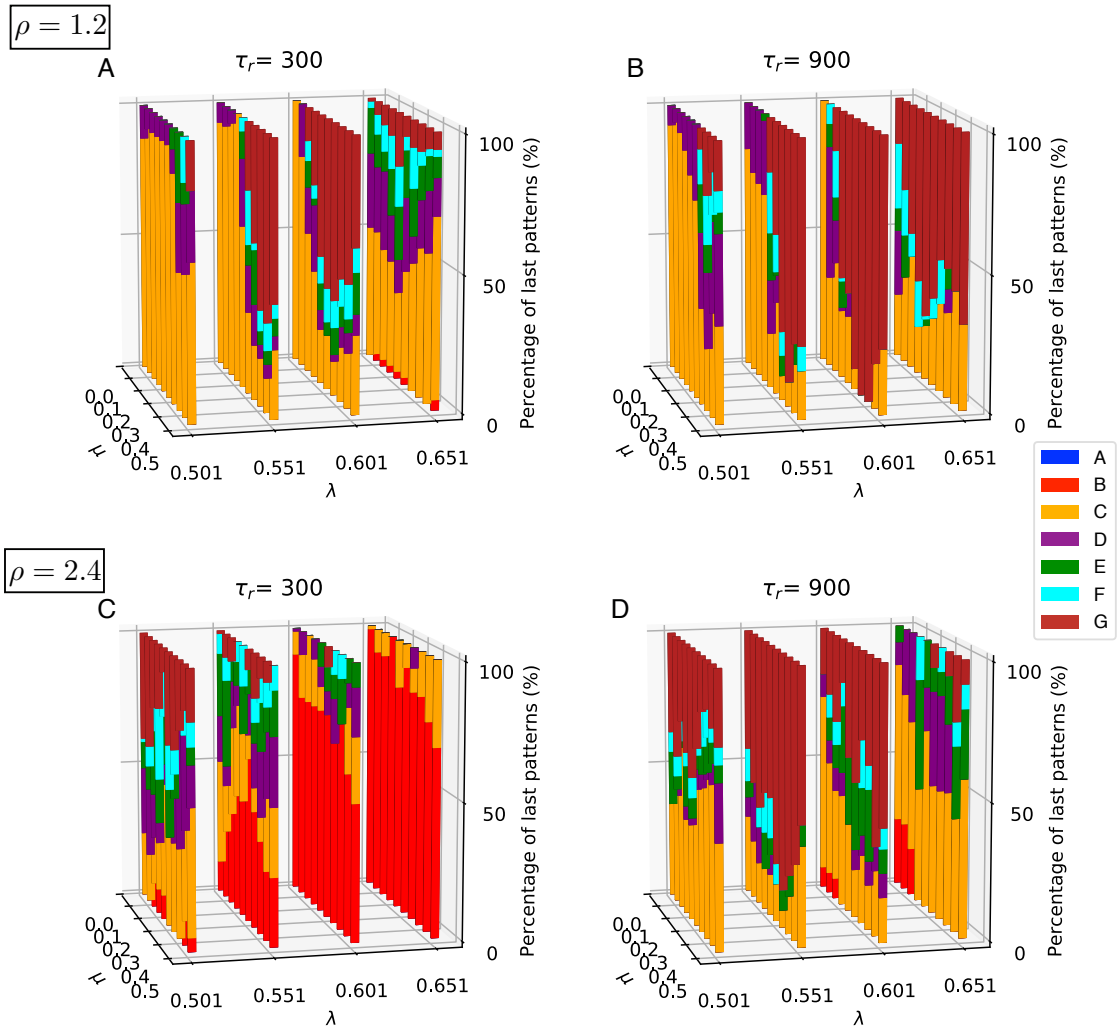

Figure 6: Percentage of last activated patterns in a regular segment over simulations for $\eta = 0.04$. Pattern colours follow to the colour codes of the last activated units (see the legend on the right). The height of each colour on a bar indicates the percentage of corresponding last activated pattern. For $\rho = 1.2$ (panels A and B), the chain length increases with τ_r. Chains are longer for intermediate values of μ, but get shorter as μ increases. The global inhibition, λ, should be high enough for the sequential activation. Activation spreads over lower values of μ as λ increases but the chain length can decrease if the inhibition is too strong. For $\rho = 2.4$ (panels C and D), the chain length increases with τ_r, but decreases with λ. Increasing μ prolongs the chains more with $\tau_r = 900$ than $\tau_r = 300$. Activation is easier with $\rho = 2.4$ than $\rho = 1.2$ if $\tau_r = 900$ and (μ, λ) are small, but the chains under strong inhibition are longer in $\rho = 1.2$ than $\rho = 2.4$.

induces an asymptotic decrease of the synaptic variable towards the value $S = (1+U\tau_r)^{-1}$. This in turn weakens the synaptic weight $J^{max}s_i$ in 1, which may destabilize ξ^i in the direction of $\hat{\xi}^i$. Considering s_i as a slowly varying parameter this can be seen as a *dynamic bifurcation* of an equilibrium along the edge from ξ^i to $\hat{\xi}^i$. The following scenario was described in [1]. For the sake of simplicity we now assume $i = 1$ (the same arguments hold for any i). The patterns ξ^1, $\hat{\xi}^1$ and ξ^2 lie at the vertices of a face, which we call F^2, generated by the coordinates x_1 and x_3, with $x_2 = 1$ and the rest of the coordinates being set to 0.
Figure 7 shows three successive snapshots of the movie on F^2. The left panel illustrates the initial configuration, with the stable pattern ξ^1 corresponding to the top left vertex. Then at some time T_0 an equilibrium bifurcates out of $\hat{\xi}^1$ in the direction of ξ^1 (here the 'slow' STD time plays the role of bifurcation parameter, see middle panel). After a time T_1 (right panel) this bifurcated equilibrium disappears in ξ^1 which becomes unstable and a connecting trajectory is created along the edge with $\hat{\xi}^1$. Simultaneously a trajectory connects $\hat{\xi}_1$ to ξ_2 on the corresponding edge. It results that the following sequence of connecting trajectories is created: $\xi^1 \to \hat{\xi}^1 \to \xi^2$. As a result, any state of the system initially close to ξ^1 will follow the 'vertical' edge towards $\hat{\xi}^1$, then the 'horizontal' edge towards ξ_2. The process can repeat itself from ξ^2 to ξ^3 and so on. It was shown in [1] that in order to work, this scenario requires that the coefficients of the matrix J^{max} satisfy the relation $J^{max}_{1,2} < J^{max}_{2,3}$ (more generally $J^{max}_{i,i+1} < J^{max}_{i+1,i+2}$, $i = 1, \ldots, P-1$, for the existence of a chain of P patterns), a condition which does not hold with (3).

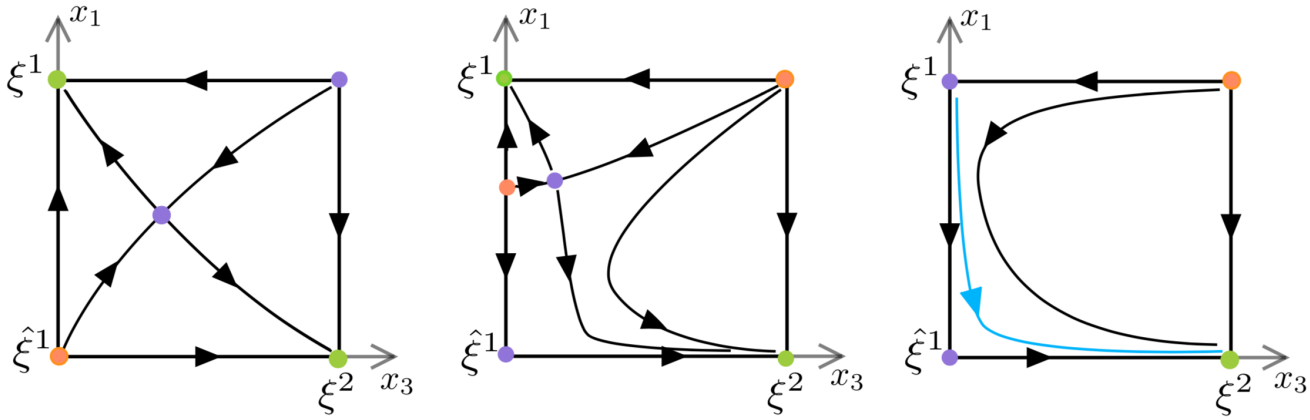

Figure 7: Phase portrait of the fast dynamics on the face F^2 at three different 'slow' STD times $t < T_0$ (the learned patterns ξ^1 and ξ^2 are stable), $T_0 < t < T_1$ (bifurcation of a saddle point on the edge between ξ^1 and $\hat{\xi}^1$) and $t > T_1$ (ξ^1 has become unstable along the edge $\xi^1 - \hat{\xi}^1$ whilst ξ^2 is still stable). Green spots are the stable equilibriums, orange spots are completely unstable and purple spots are saddles. The blue lines illustrate segments of trajectory starting near ξ^1. The saddle point in the interior of F^2 merges with the bifurcated equilibrium before right panel is realised.

To circumvent this difficulty we relax the condition that the chain of connections $\xi^1 \to \hat{\xi}^1 \to \xi^2$ exists when $t > T_1$ (right panel of Fig. 7). We assume instead that the connecting trajectory along the edge $\hat{\xi}^1 - \xi^2$ is broken by a stable equilibrium close to $\hat{\xi}^1$. In such case strong enough noise perturbations could push a state of the system which has converged to $\hat{\xi}^1$ off its basin of attraction. As a result the system would escape $\hat{\xi}^1$ and converge towards ξ^2 (in a stochastic sense), as expected. When such chains driven by noise exist, we call them *excitable chains* by reference to [2] who introduced the concept.
Under the new scheme the number of possible transitions is much larger and multiple outcomes are possible. We have identified two scenarios (named 1 and 2) by which these excitable chains can occur in our problem. Typical cases are illustrated on Figure 8. As in the previous figure, snapshots of the dynamics at three different "slow" times are shown. The red line marks the boundary of the basin of attraction of ξ^2 and the dashed circles mark the closest distances for a possible stochastic jump out of it. In both scenarios a completely unstable equilibrium point exists on the edge from ξ^1 to the unnamed vertex on F^2, which corresponds to the pattern $(1, 1, 1, 0, \ldots, 0)$ (not a learned pattern).

Under Scenario 1 the pattern ξ^1 loses first stability by a dynamic bifurcation of a sink (stable equilibrium) along the edge $\xi^1 - \hat{\xi}^1$. This sink travels along the edge until it merges with $\hat{\xi}^1$, so that a heteroclinic connection from ξ^1 to $\hat{\xi}^1$ is realised. However a saddle equilibrium bifurcates from $\hat{\xi}^1$ along the edge $\hat{\xi}^1 - \xi^2$ (middle panel). As a result $\hat{\xi}^1$ is weakly stable in the direction of ξ^2 and if noise is not too small the dynamics can jump to the basin of attraction of ξ^2 (right panel).
In Scenario 2 the picture is initially similar to that of Scenario 1. The difference comes from the simultaneous bifurcation of two equilibria from $\hat{\xi}^1$: one along the edge $\hat{\xi}^1 - \xi^1$ and the other along the edge $\hat{\xi}^1 - \xi^2$ (middle panel). The first bifurcated equilibrium eventually merges with ξ^1 so that a heteroclinic connection is created from ξ^1 to $\hat{\xi}^1$ by the same mechanism as in Figure 7 (right panel). However the other bifurcated equilibrium point prevents the formation of a heteroclinic connection from $\hat{\xi}^1$ to ξ^2. Nevertheless $\hat{\xi}^1$ is only weakly stable in the ξ^2 direction and noise can allow an incoming trajectory to jump over $\hat{\xi}^1$ towards ξ^2. Note that the equilibrium on the edge $\hat{\xi}^1 - \xi^2$ can travel towards ξ^2 as s_1 elapses, so that the basin of attraction of $\hat{\xi}^1$ becomes too large for noise to allow jumps toward ξ^2. In this case the system may get indefinitely stucked at $\hat{\xi}^1$. We refer to this behaviour as *pending*, see an example in Fig. 4, second simulation in panel (a).
It is shown in Appendix A that there indeed exist parameter domains in which one or the other of the two scenarios occurs.
Simulations (and analysis, see A.1) identify μ as the main control parameter which determines the choice between these scenarios: the system follows Scenario 1 for higher values of μ and Scenario 2 for lower values of μ. This explains the difference in behavior seen in Figure 4 at lower and higher values of μ. The boundary between the two regions is defined by the value $\mu = \mu^*$ for which ξ^i and $\hat{\xi}^i$ change stability at the same time. For an analytic definition of μ^* and more detailed analysis, see A.1.

2.4 Irregular chains

The question we address here is what happens after the last pattern of a regular segment has been reached. Here we discard the case when this last pattern remains stable indefinitely or when the system deactivates completely (all units relax to inactive state). We focus on the case when the dynamics continues afterwards by jumping to patterns in the sequence in an irregular manner and generate new chains, possibly going backward. Suppose at a time t, x_p and x_q are the two most recently activated units, with x_p preceding x_q in its activation. We define

$$\Delta = q - p.$$

Note that a regular chain satisfies $\Delta = 1$ for all t until the last pattern is reached.
We distinguished two cases of irregular continuation of chains: reversing the chain ($\Delta = -1$) and random reactivation of new chains ($|\Delta| > 1$). Recall the scenarios 1 and 2 for transitions from one pattern to the next (fig. 8). The former occurs for "large" values of μ and the latter for lower values of μ. Let $\hat{\xi}_p$ be the last intermediate state at the end of the regular segment. In Scenario 2 either $\hat{\xi}_p$ remains stable indefinitely or it destabilizes after some (long) time due to the repotentiation of s_p. The latter case corresponds to a dynamic scenario for chain's reversal. We refer to the prolonged residence of the system at $\hat{\xi}_p$ as *pending* and note that it likely leads to a reversal. However, for large values of noise, random activation ($|\Delta| > 1$) may also occur. The scenario 1 is more likely to yield random re-activation as $\hat{\xi}_p$ loses stability in the x^{p+1} direction with the decrease of s_{p+1}, so that a transition to the inactive state is possible. However in this case too other Δ values are possible when the noise is large.

Figures 9 and 10 show the percentage of Δ values after a new activation for $\eta = 0.02$ and $\eta = 0.04$, respectively. Activity with $\Delta = -1$ is generally supported for $\rho = 1.2$ and if $\rho = 2.4$, for small values of (μ, λ). As the chains get longer with increasing μ, regular segments get longer, specially when noise is high ($\eta = 0.04$). Regarding the type of forward and/or backward irregular chains, high values of ρ and high values of μ (e.g. low gain) for low values of η (Fig. 9) increase the possibility for irregular chains in the forward direction, while the combination of high values of μ (e.g. low gain), ρ and η increase the possibility for irregular chains in both directions (10).

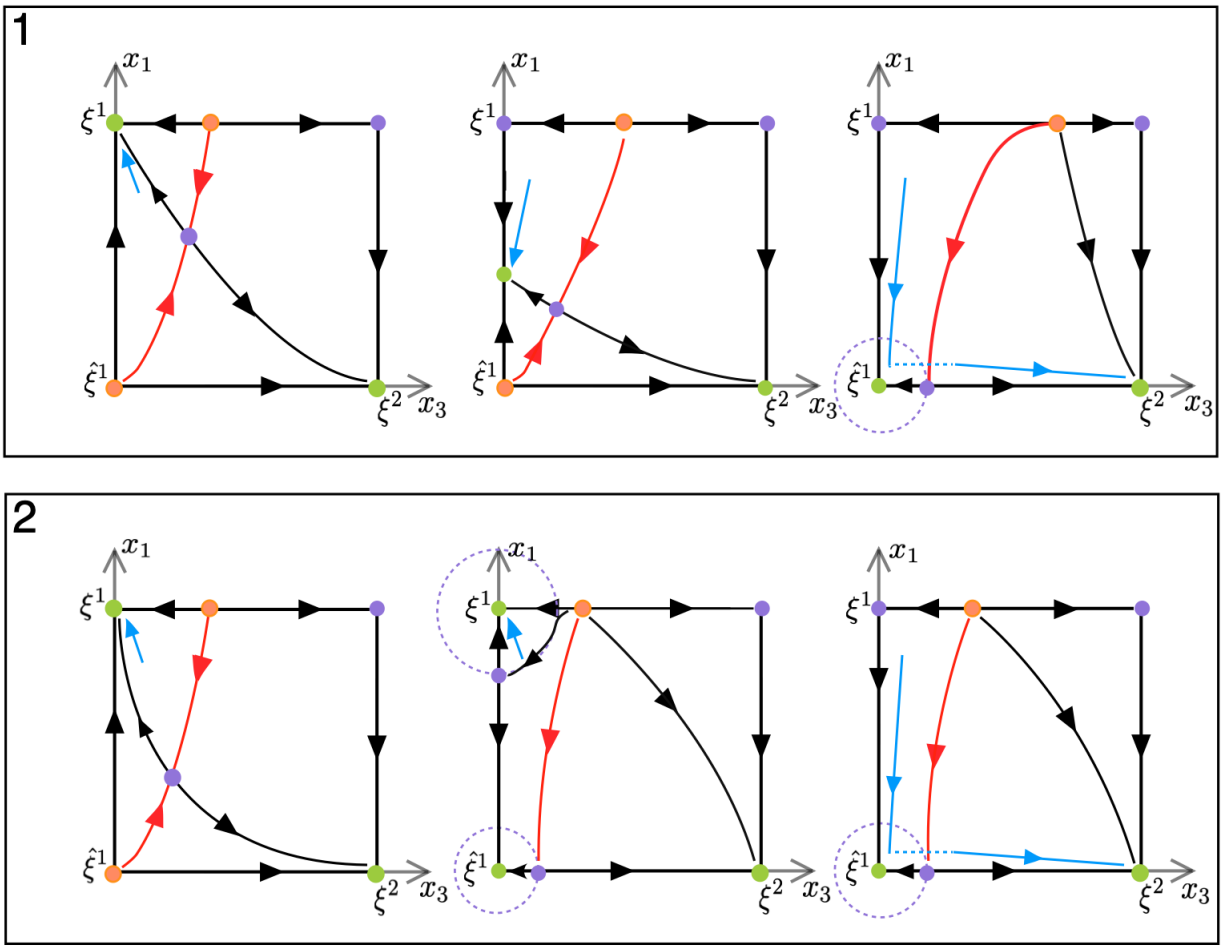

Figure 8: Phase portraits of the fast dynamics on the face F^2 at three different 'slow' STD times, illustrating the transition $\xi^1 \to \xi^2$ with excitable connections in Scenarios 1 (above) and 2 (below). Stable patterns are coloured in green. The red trajectories are separatrices between the basins of attraction of the stable equilibria. The blue lines illustrate segments of a trajectory starting near ξ^1. In panels 1 (c) and 2 (c) this trajectory "jumps" out of the basin of attraction of $\hat{\xi}^1$ under the effect of noise and converges towards ξ^2

3 Discussion

Experimental evidence indicates that the brain can either replay the same learned sequence to repeat reliable behaviors [18], [24], [25], [26], [27] or generate new sequences to create new behaviors [28], [29], [30], [31], [33] [32]. The present research identifies biologically plausible mechanisms that explain how a neural network can switch from repeting learned regular sequences to activating new irregular sequences. To make the problem analytically tractable, the combined effects of the parameters were analyzed on neuronal population firing rates in a simplified balanced network model by use of slow-fast dynamics and dynamic bifurcations. We demonstrated how variations in neuronal gain, short-term synaptic depression and noise can switch the network behavior between regular or irregular sequences for a fixed learned synaptic matrix.

3.1 Synaptic matrix

In the present model the overlap had the same number of shared units for all the overlapping populations. This allowed to show that variable overlap is not a necessary condition for the activation of sequences of populations. A consequence of the constant overlap is that sequences from a stimulus-driven end-point pattern in the sequence (e. g. first pattern A of the sequence) are directional but

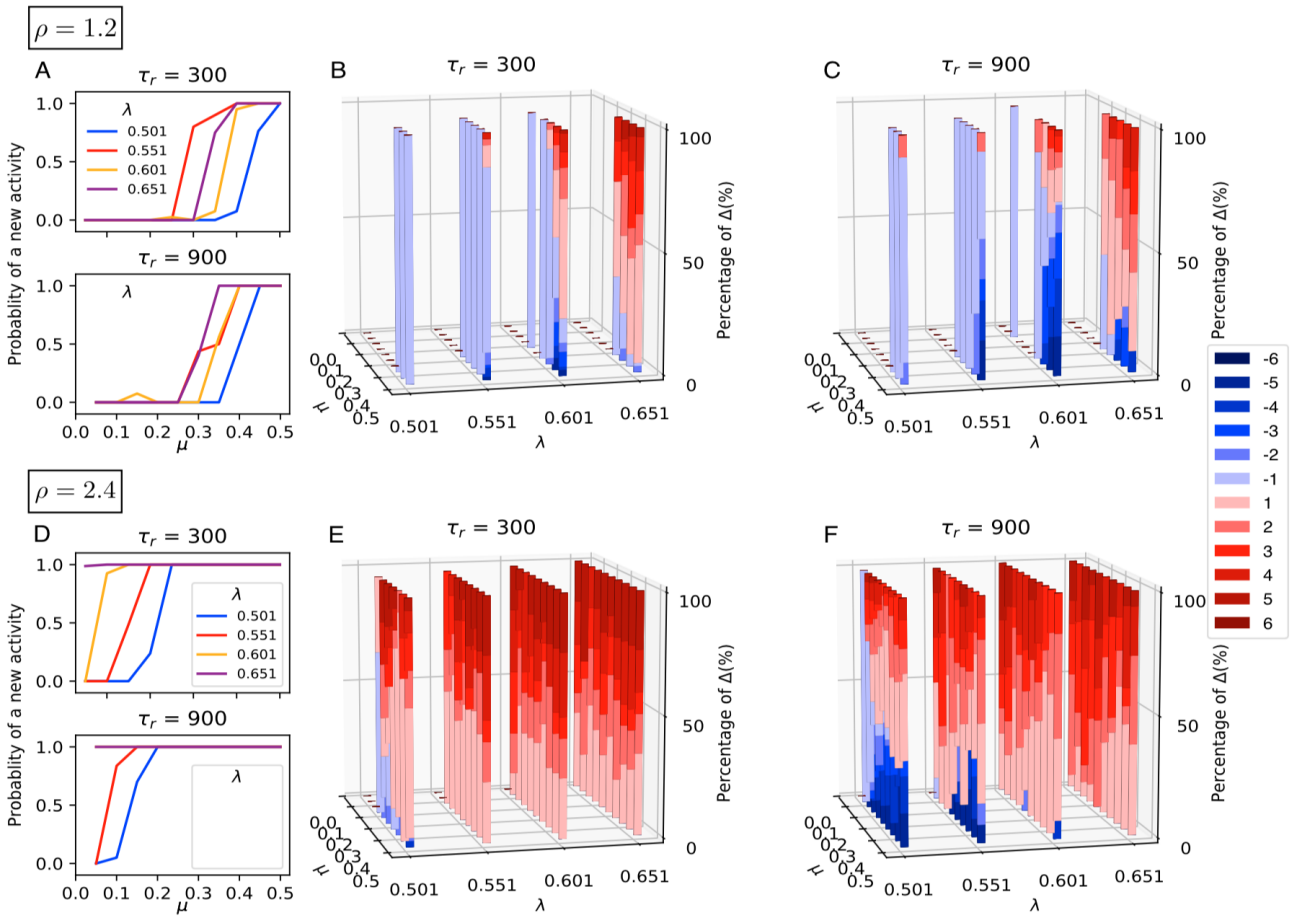

Figure 9: Activity after the initial sequence for $\eta = 0.02$. Panels A and D show the probability of a new activity to be observed after the initial sequence for $\rho = 1.2$ and $\rho = 2.4$, respectively. Bar plots in panels B, C, E, and F show the percentage of activation distance Δ if there is any activity. Bars are coloured according to the distance colours (shown on the right) and the height of each colour indicates the percentage of the corresponding distance. For $\rho = 1.2$ (panels A, B and C) and small values of μ the system tends to remain on the last activated pattern, as low rates of activity probability in panel A demonstrate. New sequences are generated as μ and λ increase, with a preference of backward activity for small values of λ. The probability of generating new sequences for small values of μ is higher if $\rho = 2.4$ and the minimum value of μ required for a new sequence decreases with λ and τ_r (panel D). For instance, if $\tau_r = 900$ and $\lambda = \{0.601, 0.651\}$, new activity occurs in all trials. The new activity starts with distance$\Delta > 0$ (panels E and F) and we observe very few cases with $\Delta < 0$ for $\rho = 2.4$ (yet, more if $\tau_r = 900$). The difference between the percentages of Δ for $\tau_r = 300$ and $\tau_r = 900$ indicates the capability of slow synapses to yield longer chains.

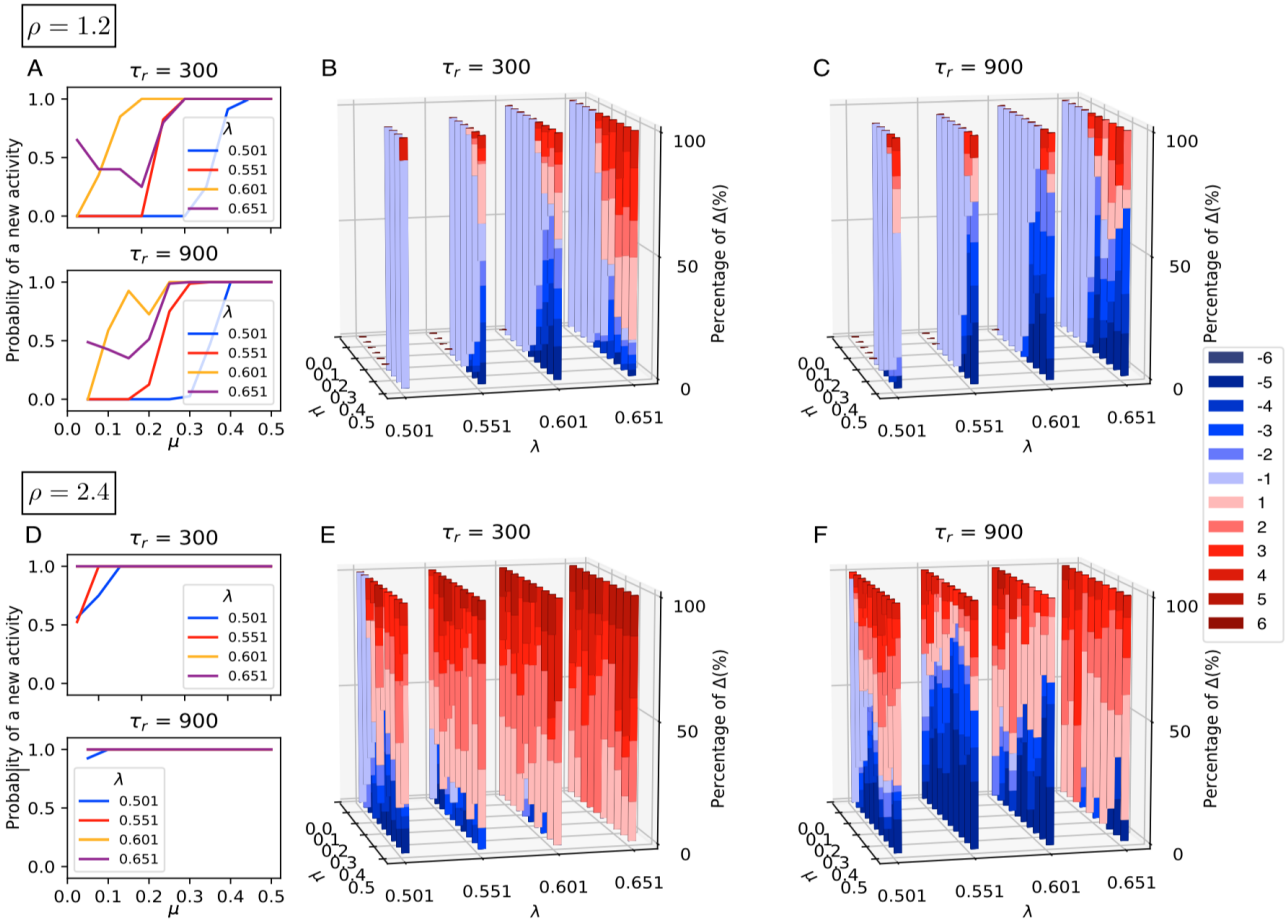

Figure 10: Activity after the initial sequence for $\eta = 0.04$. Same setting as in Fig. 9. For $\rho = 1.2$ (panels A, B and C), $\lambda = \{0.501, 0.551\}$ and small values of μ, the system remains on the last activated pattern as low rates of activity probability in panel A demonstrate. New sequences are generated as μ and λ increase, with a preference for backward activity for small values of λ. With $\rho = 2.4$ (panels D, E and F) new activation can occur for all values of μ and λ. In panels E and F, increasing λ and μ leads to an activation with $\Delta > 0$. Percentage of activity with $\Delta < 0$ is much higher with $\tau_r = 900$ than $\tau_r = 300$. The difference indicates the capability of slow synapses to yield longer chains both for $\rho = 1.2$ and $\rho = 2.4$.

sequences from a mid-point pattern can go in any of the two possible directions. The model can then generate bi-directional sequences interesting in free recall. Starting from the first pattern A (or G), the sequence ABCDEFG is oriented in one direction (or in the other direction), and starting from the first pattern e.g. D, the sequence can be oriented in any of the two possible directions. The present model allows for bi-directionnal sequences as well as for new sequences depending on the value of neuronal gain $\gamma = \mu^{-1}$.

3.2 Regular *vs.* irregular sequences

Regarding regular sequences, the length of the chain activated increases with noise and for combinations of strong STD (high values of ρ) and low inhibition, or weak STD (low values of ρ) and strong inhibition. Further, for most combinations of noise, STD and inhibition, there is an optimal value of gain that generates the longest chains. The sensitivity of a neuron to its incoming activation varies with changes in its gain [10]. Simulations and analysis show that the neuronal gain is a key control parameter that selects the length and type of sequence activated: regular or iregular. Large neuronal gain impairs the deactivation of the units in a pattern and hence makes the transition to the next pattern difficult, and small gain impairs the activation of the next unit and again makes the transition difficult. Consequently there is an optimal window for the gain corresponding to long sequences. Experimental evidence shows that presentations of a given stimulus reproduces the same sequence reliably [26] [27]; [62] [34]. The present model can repeat systematic full sequences of activation for some values of the parameters that make the network change patterns in a given order. This 'reliable' mode could be well adapted to the reliable reproduction of learned sequences of behaviors.

Regarding irregular sequences, a large neuronal gain leads to the second scenario describing transitions from one pattern to the next (as in fig. 8). According to this second scenario, the last intermediate state of the network at the end of a regular segment can destabilize and leads to a reversal of the sequence. In that case the network activates patterns backward in the reverse order. Further, for large values of noise, scenario two can lead to random activation of patterns in either the forward or backward direction. Such variable sequences are more likely to be generated according to the first scenario that makes possible a transition to another state in the forward or backward direction and that does not necessarily overlap with the current state (forward or backward leaps). Our model can generate variable sequences over repetitions of a same triggering stimulus for high values of gain, in line with a *memoryless* system [2] that activates a new pattern in an unpredictable fashion. Behavioral studies indicate that presentation of a triggering stimulus can activate distant items that are not directly associated to it [39]. The generation of new sequences corresponding to the activation of new possibilities [15] and the execution of new information-seeking behaviors such as saccades or locomotor explorations of unknown locations [14]. This 'creative' mode of variable activation not following a given sequence could correspond to a mind wandering mode [29] [65] or divergent thinking involved in creativity [68] [67] [64] [63].

3.3 Neuromodulation of the switch between regular and irregular sequences

A novel feature of our network model is that neuronal gain influences the type of sequences that are generated: regular or irregular. Typical computational models of sequence generation reproduce learned sequences [18]. However, if the brain must in some case reproduce systematic behaviors, it must also have the capacity to liberate itself from repetition in order to create new behaviors. The present research shows that the network can exhibit the dual behavior of activating regular or irregular sequences for a given synaptic matrix. The transition depends on biological parameters, in particular on gain modulation. Given that changes in gain change the length of the regular sequence, and that when the regular sequence stops it becomes irregular, the gain controls the regularity of the sequences. The gain is reported to depend on neuromodulatory factors such as dopamine [85]

[72][70] [71] involved in reward-seeking behaviors and punishment [69] [73] [74]. Dopamine is reported to modulate the magnitude of the activation between associates in memory (priming; [37] [38]) and dopamine induced changes in neuronal gain have been reported to account for changes in activation in memory [40] and for movement control [86]. The present research sheds light on how the brain can switch between a 'reliable' mode and a 'creative' mode of sequencial behavior depending on external factors such as reward that neuromodulate neuronal gain.

Acknowledgements

EKE and MK were supported by the ERC Advanced Grant NerVi no. 227747.
The authors thank Gianluigi Mongillo for helpful discussions.

References

[1] Aguilar C., Chossat P., Krupa M. and Lavigne F. (2017). Latching dynamics in neural networks with synaptic depression. *PLoS One*, 2 (8), e0183710. https://doi.org/10.1371/journal.pone.0183710

[2] Ashwin P. and Postlethwaite C. (2016). Designing heteroclinic and excitable networks in phase space using two populations of coupled cells. *J Nonlinear Sci* 26: 345. https://doi.org/10.1007/s00332-015-9277-2

[3] Lerner I., Bentin S. and Shriki O. (2012a). Spreading activation in an attractor network with latching dynamics: automatic semantic priming revisited. *Cogn. Sci.*, **36**, 1339 –1382. doi:10.1111/cogs.12007.

[4] Lerner I. and Shriki O. (2014). Internally and externally driven network transitions as a basis for automatic and strategic processes in semantic priming: theory and experimental validation. *Front. Psychol.* **5:314**. doi:10.3389/fpsyg.2014.00314.

[5] Treves A. (2005). Frontal latching networks: a possible neural basis for infinite recursion. *Cognitive Neuropsych.* **22**(3-4): 276–291.

[6] Tsodyks M., Pawelzik K. and Markram H. (1998). Neural networks with dynamic synapses. *Neural Computation*, 10, 821-835. https://doi.org/10.1162/089976698300017502

[7] Romani S., Pinkoviezky I., Rubin A. and Tsodyks M. (2013) Scaling laws of assosciative memory retrieval. *Neural Computation*, **25**: 2523-2544. https://doi.org/10.1162/NECO_a_00499.

[8] Katkov M., Romani S., and Tsodyks M. (2017) Memory retrieval from first principles. *Neuron*, **94**: 1027–1032. http://dx.doi.org/10.1016/j.neuron.2017.03.048

[9] Salinas E., and Thier, P. (2000). Gain modulation: a major computational principle of the central nervous system. Neuron 27, 15-21.

[10] Salinas E., and Sejnowski, T. J. (2001). Gain modulation in the central nervous system: where behavior, neurophysiology, and computation meet. Neuroscientist 7, 430-440.

[11] Silver, A. (2010), Neuronal arithmetic. Nature Reviews Neuroscience 11, 474 – 489.

[12] Brunel, N., and Lavigne, F. (2009) Semantic priming in a cortical network model. em J. Cog. Neurosci., **21**: 2300-2319.

[13] Lavigne, F., Dumercy, L. & Darmon, N. (2011) Determinants of Multiple Semantic Priming: A Meta-Analysis and Spike Frequency Adaptive Model of a Cortical Network. *The Journal of Cognitive Neuroscience*, **23(6)**: 1447-1474.

[14] Pezzulo, G., van der Meer, M. A. A., Lansink, C. S., & Pennartz C. M. A. (2014) Internally generated sequences in learning and executing goal-directed behavior. *Trends in Cognitive Sciences*, **18(12)**: 647-657.

[15] Hassabis, D. et al. (2007) Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 1726–1731.

[16] Rohrmeier, M. A. & Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *International Journal of Psychophysiology*, **83**: 164-175.

[17] Bubic A., Von Cramon, D. Y. & Schubotz, R. (2010). Prediction, cognition and the brain (2010). *Frontiers in Human Neuroscience*, DOI=10.3389/fnhum.2010.00025.

[18] Veliz-Cuba, A., Shouval, H. Z., Josic, K. & Kilpatrick, Z. P. (2015). Networks that learn the precise timing of event sequences. *J Comput Neurosci*, **39**: 235-254.

[19] Burgess, N. & Hitch, G.J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, **106(3)**: 551.

[20] Zatorre, R.J., Chen, J.L. & Penhune, V.B. (2007). When the brain plays music: auditorymotor interactions in music perception and production. *Nature Reviews Neuroscience*, **8(7)**: 547–558.

[21] Graziano, M., Polosecki, P., Shalom, D.E. & Sigman, M. (2011). Parsing a perceptual decision into a sequence of moments of thought. *Frontiers in Integrative Neuroscience*, DOI=10.3389/fnint.2011.00045.

[22] Meyer, T. & Olson, C.R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences*, **108(48)**: 19,401–406.

[23] Kok, P., Jehee, J.F. & de Lange F.P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, **75(2)**: 265–270.

[24] Conway, C.M. & Christiansen, M.H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, **5(12)**: 539–546.

[25] Buhusi, C.V. & Meck, W.H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, **6(10)**: 755–765.

[26] Eagleman, S.L. & Dragoi, V. (2012). Image sequence reactivation in awake v4 networks. *Proceedings of the National Academy of Sciences of the United States of America*, **109(47)**: 19, 450–455.

[27] Xu, S., Jiang, W., Poo, Mm. & Dan, Y. (2012). Activity recall in a visual cortical ensemble. *Nature Neuroscience*, **15**: 449–455.

[28] R.L. Buckner, J.R. Andrews-Hanna & D.L. Schacter (2008). The brain's default network *Ann. NY Acad. Sci.*, **1124**: 1-38.

[29] Christoff, A.M. Gordon, J. Smallwood, R. Smith, J.W. Schooler (2009). Experience sampling during fMRI reveals default network and executive system contribution to mind wandering. *Proceedings of the National Academy of Sciences, USA*, **106**: 8719-8724.

[30] Guilford, J.P. (1950). Creativity. *American Psychologist*, **5**: 444-454.

[31] Abraham, A., Beudt, S., Ott, D.V. M. & von Cramon, D.R. (2012). Creative cognition and the brain: Dissociations between frontal, parietal-temporal and basal ganglia groups. *Brain Research*, **1482**: 55-70.

[32] Fink, A. & Benedek, M. (2012). EEG alpha power and creative ideation. *Neuroscience and Biobehavioral Reviews*, **44**: 11-123.

[33] Gonen-Yaacovi, G., de Souza, L.C., Levy, R., Urbanski, M. Josse, G. & Volle Emmanulle. (2013). Rostral and caudal prefrontal contributions to creativity: a meta-analysis of functional imaging data. *Frontiers in Human Neuroscience*, **7**: 465.

[34] Gavornik, J.P. & Bear, M.F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nature Neuroscience*, **17(5)**: 732–737.

[35] Jenkins, I., Brooks, D., Nixon, P., Frackowiak, R. & Passingham, R. (1994). Motor sequence learning: a study with positron emission tomography. *The Journal of Neuroscience*, **14(6)**: 3775–3790.

[36] Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Sasaki, Y. & Ptz, B. (1998). Transition of brain activation from frontal to parietal areas in visuomotor sequence learning. *The Journal of Neuroscience*, **18(5)**: 1827–1840.

[37] Roesch-Ely, D.,Weiland, S., Scheffel, H., Schwaninger,M., Hundemer, H.-P., Kolter, T., et al. (2006). Dopaminergic modulation of semantic priming in healthy volunteers. *Biological Psychiatry*, **60(6)**, 604611.

[38] Kischka, U., Kammer, T. H., Maier, S.,Weisbrod, M., Thimm, M., & Spitzer, M. (1996). Dopaminergic modulation of semantic network activation. *Neuropsychologia*, **34(11)**, 11071113.

[39] Bowden, E. M., & Jung-Beeman, M. (2003). One hundred forty-four Compound Remote Associate Problems: Short insight-like problems with one-word solutions. *Behavioral Research, Methods, Instruments, and Computers*, **35**, 634-639.

[40] Lavigne, F. & Darmon, N. (2008). Dopaminergic Neuromodulation of Semantic Priming in a Cortical Network Model. *Neuropsychologia*. **46**, 3074-3087.

[41] Hung, C., Kreiman, G., Poggio,T., DiCarlo, J. (2005). Fast read-out of object information in inferior temporal cortex. *Science*, **310**: 863866.

[42] Young, M.,Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science*, **256**: 13271331.

[43] Kreiman, G., Hung, C. P., Kraskov, A.,Quian Quiroga, R.,Poggio,T. and DiCarlo, J. J. (2006). Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. *Neuron*, **49**: 433445.

[44] Quian Quiroga, R. & Kreiman, G. (2010). Measuring sparseness in the brain: comment on Bowers (2009). *Psychol. Rev.*, **117**: 291299.

[45] Quian Quiroga R. (2016). Neuronal codes for visual perception and memory. *Neuropsychologia*, **83**: 227-41. doi: 10.1016/j.neuropsychologia.2015.12.016.

[46] Miyashita, Y. (1988) Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, **335**: 817-820.

[47] Erickson, C. A., and Desimone, R. (1999) Responses of macaque perirhinal neurons during and after visual stimulus association learning. *J. Neurosci.*, **19**: 10404-10416.

[48] Rainer, G., Rao, S. C., and Miller, E. K. (1999) Prospective coding for objects in primate prefrontal cortex. *J. Neurosci.*, **19**: 5493-5505.

[49] Reddy, L., Poncet, M., Self, M., W., Peters, J. C., Douw, L. van Dellen, E., Claus, S., Reijneveld, J., C., Baayen, J. C., and Roelfsema, P. R. (2015). Learning of anticipatory responses in single neurons of the human medial temporal lobe. *Nature comm.*, **6:8556**: DOI: 10.1038.

[50] Yakovlev V, Fusi S, Berman E, Zohary E. (1998). Inter-trial neuronal activity in inferior temporal cortex: a putative vehicle to generate long-term visual associations. *Nat Neurosci.* **1(4)**:310-7.

[51] Weinberger N. M. (1998). Physiological memory in primary auditory cortex: characteristics and mechanisms. *Neurobiol Learn Mem*, **70(1-2)**:226-51.

[52] Mongillo, G., Amit, D. J., & Brunel, N. (2003). Retrospective and prospective persistent activity induced by Hebbian learning in a recurrent cortical network. *Eur J Neurosci*, **18(7)**: 2011-2024.

[53] Brunel, N. (1996). Hebbian learning of context in recurrent neural networks. *Neural Comput.*, **15**(8):1677-710.

[54] Lavigne F, Denis S (2001) Attentional and semantic anticipations in recurrent neural networks. *Int J Comput Anticip Syst*, **14**:196214.

[55] Lavigne F, Denis S (2002) Neural network modeling of learning of contextual constraints on adaptive anticipations. *Int J Comput Anticip Syst*, **12**:253268.

[56] Rolls E. T., Tovee M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J Neurophysiol.* **73(2)**:713-26.

[57] Tamura,H.,Tanaka,K. (2001). Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cereb. Cortex.* **11**: 384-399.

[58] Tsao, D. Y., Freiwald, W. A., Tootell, R. B.,Livingstone, M. (2006). A cortical region consisting entirely of face-selective cells. *Science*, **311**: 670-674.

[59] Fujimichi R., Naya Y., Koyano K. W., Takeda M., Takeuchi D. and Miyashita Y. (2010). Unitized representation of paired objects in area 35 of the macaque perirhinal cortex. *Eur J Neurosci.*, **32(4)**: 659-67. doi: 10.1111/j.1460-9568.2010.07320.x.

[60] Quian Quiroga, R. (2012). Concept cells: the building blocks of declarative memory functions. *Nat. Rev. Neurosci.*, **13**: 587-597.

[61] Tsodyks, M., V. 1990. Associative memory with binary synapses. *Modern Physics Letters B* **11**: 713-716.

[62] Shuler, M., G, and Bear, M., F. (2006). Reward timing in the primary visual cortex. *Science*, **311**: 1606-1609.

[63] Benedek, M., and Neubauer, A., C. (2013). Revisiting Mednick's model on creativityrelated differences in associative hierarchies. Evidence for a common path to uncommon thought. *The Journal of creative behavior*, **47(4)**: 273-289.

[64] Benedek, M., Knen, T., and Neubauer, A., C. (2012). Associative abilities underlying creativity. *Psychology of Aesthetics, Creativity, and the Arts*, **6(3)**: 273.

[65] Andrews-Hanna, J., R. (2012). The brains default network and its adaptive role in internal mentation. *The Neuroscientist*, **18(3)**: 251-270.

[66] Christoff, K., Gordon, A., M., Smallwood, J., Smith, R., & Schooler, J., W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences*, **106(21)**: 8719-8724.

[67] Guilford, J. P. (1967). *The nature of human intelligence.*

[68] Beaty, R. E., Benedek, M., Silvia, P. J. & Schacter, D. L. (2016). Creative cognition and brain network dynamics. *Trends in cognitive sciences*, **20(2)**: 87-95.

[69] Jhou, T. C. & Vento, P. J. (2019). Bidirectional regulation of reward, punishment, and arousal by dopamine, the lateral habenula and the rostromedial tegmentum (RMTg). *Current Opinion in Behavioral Sciences*, **26**: 90-96.

[70] Braver, T. S., Barch, D. M. & Cohen, J. D. Cognition and control in schizophrenia: a computational model of dopamine and prefrontal function. *Biological Psychiatry*, **46**: 312-328.

[71] Cohen, J. D. & Servan-Schreiber, D. Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, **99**: 45-77.

[72] Seamans, J. K., Durstewitz, D., Christie, B. R., Stevens, C. F. and Sejnowski, T. J. Dopamine D1/D5 receptor modulation of excitatory synaptic inputs to layer V prefrontal cortex neurons (2001). *Proceedings of the National Academy of Sciences*, **98**: 301-306.

[73] Lak, A., Stauffer, W. R., & Schultz, W. (2014). Dopamine prediction error responses integrate subjective value from different reward dimensions. *Proceedings of the National Academy of Sciences*, **111(6)**: 2343-2348.

[74] Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, **442(7106)**: 1042.

[75] Burgess, N., and Hitch, G. J. (1999). Memory for serial order: a network model of the phonological loop and its timing. *Psychological review*, **106(3)**, 551.

[76] Bick, C., and Rabinovich, M. I. (2009). Dynamical origin of the effective storage capacity in the brains working memory. *Physical Review Letters*, **103(21)**, 218101.

[77] Tsodyks, M. V., and Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proceedings of the national academy of sciences*, **94(2)**, 719-723.

[78] Huber, D. E., and O'Reilly, R. C. (2003). Persistence and accommodation in shortterm priming and other perceptual paradigms: temporal segregation through synaptic depression. *Cognitive Science*, **27(3)**, 403-430.

[79] Servan-Schreiber, D., Printz, H., and Cohen, J. D. (1990). A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science*, **249(4971)**, 892-895.

[80] Aston-Jones, G., and Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, **28**, 403-450.

[81] Kandel, E. R. (2001). The molecular biology of memory storage: a dialogue between genes and synapses. *Science*, **294(5544)**, 1030-1038.

[82] Alberini, C. M. (2009). Transcription factors in long-term memory and synaptic plasticity. *Physiological reviews*, **89(1)**, 121-145.

[83] Takeuchi, T., Duszkiewicz, A. J., and Morris, R. G. (2014). The synaptic plasticity and memory hypothesis: encoding, storage and persistence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369(1633)**, 20130288.

[84] Nabavi, S., Fox, R., Proulx, C. D., Lin, J. Y., Tsien, R. Y., and Malinow, R. (2014). Engineering a memory with LTD and LTP. *Nature*, **511(7509)**, 348.

[85] Rolls, E. T., Loh, M., Deco, G., and Winterer, G. (2008). Computational models of schizophrenia and dopamine modulation in the prefrontal cortex. *Nature Reviews Neuroscience*, **9(9)**, 696.

[86] Stroud, J. P., Porter, M. A., Hennequin, G., & Vogels, T. P. (2018). Motor primitives in space and time via targeted gain modulation in cortical networks. *Nature neuroscience*, **21(12)**, 1774.

A Latching dynamics from the slow-fast view point

Throughout this work we have assumed that the firing rates x_i evolve on a faster time scale than the synaptic variables s_i. This can be formalized by redefining (1) as a slow-fast system:

$$\begin{aligned} \dot{x_i} &= x_i(1-x_i)\left(-\mu x_i - I - \lambda\sum_{j=1}^{N} x_j + \sum_{j=1}^{N} J_{i,j}^{max} s_j x_j\right) + \eta \\ \dot{s}_i &= \varepsilon((1-s_i) - \rho s_i x_i) \end{aligned} \tag{6}$$

where $\varepsilon = 1/\tau_r$ and $\rho = \tau_r U$ is the parameter introduced in Section 2.1. To keep our presentation consistent with [1] we use a slightly more general version of the model, adding the parameter I that can be understood as feedforward inhibition or as modulation of the excitability of the ith unit. Using the formulation (6) we can apply the tools of the slow-fast systems' theory to gain insight into the transitions between learned patterns. Here we will make some basic observations, leaving rigorous slow-fast analysis for a later publication.

A.1 Fast subsystem dynamics in the 2D transition plane

By setting $\varepsilon = 0$ in (6), we obtain the fast subsystem of (6) where s_i's are considered as parameters. This underlies the idea of dynamic bifurcation: as the s_i's change, the features of the dynamics of the fast system, in particular the stability properties of the patterns ξ_i, evolve. We keep in mind, however, that the s_i's must follow the slow flow, so their values are not arbitrary and are related to each other. We are particularly interested in transitions from ξ_i to ξ_{i+1}, with the dynamics passing near $\hat{\xi}_i$. During this transition x_i moves from near 1 to near 0, x_{i+1} stays close to 1, and x_{i+2} moves from near 0 to near 1. The remaining units stay near 0. Hence the relevant dynamics is approximated by the restriction of the fast system to the plane :

$$F^{i+1} = \{(x_1, \ldots, x_N) \ : \ x_{i+1} = 1, \ x_j = 0 \text{ if } j \neq \{i, i+2\}\},$$

given by:

$$\begin{aligned} \dot{x}_i &= x_i(1-x_i)\left(-\mu x_i - I - \lambda(1 + x_i + x_{i+2}) + 2s_i x_i + s_{i+1}\right) \\ \dot{x}_{i+2} &= x_{i+2}(1-x_{i+2})\left(-\mu x_{i+2} - I - \lambda(1 + x_i + x_{i+2}) + 2s_{i+2} x_{i+2} + s_{i+1}\right). \end{aligned} \tag{7}$$

Since $0 \leq x_i \leq 1$ we are interested in the dynamics of (7) restricted to the square $[0,1]^2$, whose edges are invariant for the dynamics. The equilibrium points $(x_{i+2}, x_i) = (0,1)$ and $(x_{i+2}, x_i) = (1,0)$ represent ξ^i and ξ^{i+1}, respectively, while $(x_{i+2}, x_i) = (0,0)$ represents $\hat{\xi}^i$ (the relevant picture is given in Fig 8, with the subscripts 1 and 2 replaced by i and $i+1$).

A.2 Dynamic bifurcation scenarios

As s_i and s_{i+1} vary in $[S,1]$ (see Section 2.1 for the definition of S), the equilibria of (7) undergo several bifurcations that are responsible for the transition $\xi^i \to \hat{\xi}^i \to \xi^{i+1}$, which, in the particular context of (7), occurs when the eigenvalues of ξ_i or $\hat{\xi}_i$ corresponding to the edge $x_{i+2} = 0$ pass through 0. These eigenvalues can be easily read off from (7):

$$\begin{aligned} \lambda_i &= -(\mu + 2\lambda + I) + 2s_i + s_{i+1} \\ \hat{\lambda}_i &= -(\lambda + I) - s_{i+1}. \end{aligned} \tag{8}$$

Note that both λ_i and $\hat{\lambda}_i$ are time dependent and vary on the same time scale as the s's. According to the above we are interested in the following bifurcations:

- The initially stable equilibrium point $(x_{i+2}, x_i) = (0,1)$ becomes unstable at

  $$2s_i(t_{(0,1)}) + s_{i+1}(t_{(0,1)}) = \mu + 2\lambda + I,$$

  where $t_{(0,1)}$ is the corresponding bifurcation moment.

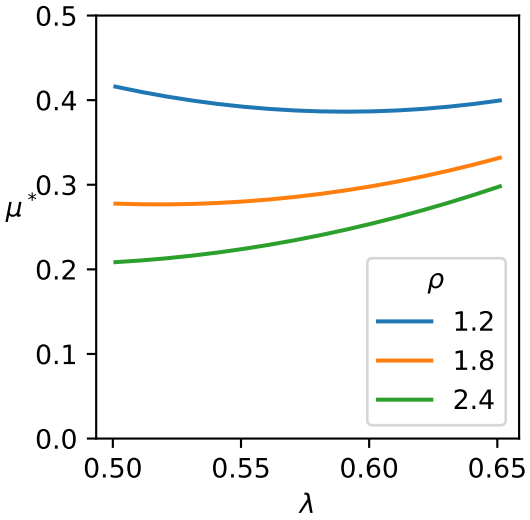

Figure 11: Variation of μ^* versus λ for $\rho = \{1.2, 1.8, 2.4\}$. The function $\mu^*(\lambda)$ changes non-monotonically for $\rho = \{1.2, 1.8\}$ with minima at $(\lambda, \mu) = (0.591, 0.3863)$ and $(\lambda, \mu) = (0.521, 0.2768)$, respectively, but increases monotonically for $\rho = 2.4$.

- Initially unstable equilibrium point $(x_{i+2}, x_i) = (0, 0)$ becomes stable at

$$s_{i+1}(t_{(0,0)}) = \lambda + I,$$

  where $t_{(0,0)}$ is the corresponding bifurcation moment.

The order of these bifurcations depends on the values of μ, λ and I and we identified two different dynamic bifurcation scenarios: Scenario 1 if $t_{(0,1)} < t_{(0,0)}$ and Scenario 2 if $t_{(0,0)} < t_{(0,1)}$.
The phase portraits defining Scenario 1 are sketched in panel 1 of Figure 8 and can be characterized as follows:

- Phase portrait (a) in panel 1 corresponds to the situation where $\lambda < 0$ and $(0, 1)$ is stable. At the same time $\hat{\lambda} > 0$ and $(0, 0)$ is unstable.
- Phase portrait (b) in panel 1 corresponds to the case $\lambda_i > 0$ and $\hat{\lambda}_i > 0$. In this case $(0, 1)$ has become unstable and there exists a stable equilibrium on the x_i-axis, $(x_{i+2}, x_i) = (0, x_i^*)$. The point $(0, 0)$ is still unstable.
- $\lambda_i > 0$ and $\hat{\lambda}_i < 0$. In this case $(0, 0)$ becomes stable and there exists a saddle at $(x_{i+2}^*, 0)$. There is a possibility of an *excitable* connection, when trajectories travel down to $(0, 0)$, jump over to $(x_{i+2}^*, 0)$ by the action of noise, and continue on to $(1, 0)$.

The phase portraits defining Scenario 2 are sketched in panel 2 of Figure 8 and can be characterized as follows:

- Phase portrait (a) in panel 2 corresponds to the situation where $\lambda_i < 0$ and $(0, 1)$ is stable. At the same time $\hat{\lambda} > 0$ and $(0, 0)$ is unstable.
- Phase portrait (b) in panel 2 corresponds to the case when both λ_i and $\hat{\lambda}_i$ are negative. In this case $(0, 1)$ is still stable, while $(0, 0)$ has become also stable and there exists a saddle type equilibrium on the x_i-axis, $(x_{i+2}, x_i) = (0, x_i^*)$.
- Phase portrait (c) in panel 2 corresponds to the case $\lambda_i > 0$ and $\hat{\lambda}_i < 0$. In this case $(0, 1)$ has become unstable and trajectories can pass from $(1, 0)$ to $(0, 0)$ and jump over a saddle point on the line $x_i = 0$, making an excitable connection to $(0, 1)$.

The feature that distinguishes Scenarios 1 and 2 is that the saddle $(0, x^*_{i+2})$ present in Scenario 1 bifurcates dynamically from $(0,0)$. Hence the jump required is arbitrarily small.

Near the bifurcation of $(0,0)$ at $\hat{\lambda}_i = 0$, $s_{i+1} = I + \lambda$ and $s_{i+2} \approx 1$, whereas the value of s_i is harder to compute, but we know that

$$S < s_i < s_{i+1} = I + \lambda.$$

Note that when $\hat{\lambda} = 0$ (equivalently $s_{i+1} = \lambda + I$), the equilibrium at $(0,0)$, corresponding to $\hat{\xi}^i$ for the original system, has two 0 eigenvalues.

We now determine $\mu = \mu^*$ which separates Scenario 1 from Scenario 2. Note that μ^* is defined by the requirement $t_{(0,0)} = t_{(1,0)}$. We will count the time t starting with the previous transition $\hat{\xi}_{i-1} \to \xi_i$ and we will assume that this transition is instantaneous. Note that at that time $s_{i+1} \approx 1$ (this holds for any μ). Given that $s_{i+1}(t_{(0,0)}) = \lambda + I$ we obtain, using the slow equation:

$$t_{(0,0)} = \int_1^{\lambda+I} \frac{ds}{\varepsilon(1 - s(1+\rho))}.$$

Using the assumptions $t_{(0,0)} = t_{(1,0)}$ and $\hat{\lambda}_i(t_{(1,0)}) = 0$ we obtain $2s_i(t_{(0,0)}) = \lambda + \mu$. Given that the computation of s_i is independent of i we now that $s_i = \lambda + I$ at the time of the transition $\hat{\xi}_{i-1} \to \xi_i$. Hence

$$t_{(0,0)} = t_{(1,0)} = \int_{\lambda+I}^{\frac{\mu+\lambda}{2}} \frac{ds}{\varepsilon(1 - s(1+\rho))}$$

This implies that μ^* is given by the following equation:

$$\int_1^{\lambda+I} \frac{ds}{\varepsilon(1 - s(1+\rho))} = \int_{\lambda+I}^{\frac{\mu^*+\lambda}{2}} \frac{ds}{\varepsilon(1 - s(1+\rho))} \tag{9}$$

Decreasing μ from μ^* gives $t_{(1,0)} > t_{(0,0)}$, i.e. Scenario 2 and increasing μ from μ^* gives $t_{(0,0)} > t_{(1,0)}$, i.e. Scenario 1.

There are some additional constraints that are shared between the two scenarios. First, we require that $\hat{\xi}^i$ should be stable in the transverse directions, in particular in the direction of x_{i+1}. The relevant eigenvalue is $\hat{\sigma}_{i+1} = \mu + I + \lambda - 2s_{i+1}$, which gives, upon substitution of $s_{i+1} = I + \lambda$, $\hat{\sigma}_{i+1} = \mu - I - \lambda$. This introduces another condition:

$$\mu < \lambda + I. \tag{10}$$

Second, as in [1], we require the stability of ξ^i in the absence of synaptic depression and the stability of $\hat{\xi}^i$ in transverse directions. This implies:

$$I + 2\lambda + \mu < 2, \quad I + \lambda < 1 < I + 2\lambda. \tag{11}$$

Figure 11 shows the μ^* decreasing with ρ: μ^* increases with λ for $\rho = 2.4$, whereas its minimum is in the middle ranges of λ for $\rho = 1.2$. The form of the $\mu^*(\lambda)$ function explains why the system gives longer chains under weak inhibition for $\rho = 2.4$, but middle/strong inhibition for $\rho = 1.2$ in Figures 5 and 6.