OPEN ACCESS Freely available online

PLOS ONE

Nonlinear Dynamics Analysis of a Self-Organizing Recurrent Neural Network: Chaos Waning

Jürgen Eser, Pengsheng Zheng, Jochen Triesch*

Frankfurt Institute for Advanced Studies, Frankfurt am Main, Germany

Abstract

Self-organization is thought to play an important role in structuring nervous systems. It frequently arises as a consequence of plasticity mechanisms in neural networks: connectivity determines network dynamics which in turn feed back on network structure through various forms of plasticity. Recently, self-organizing recurrent neural network models (SORNs) have been shown to learn non-trivial structure in their inputs and to reproduce the experimentally observed statistics and fluctuations of synaptic connection strengths in cortex and hippocampus. However, the dynamics in these networks and how they change with network evolution are still poorly understood. Here we investigate the degree of chaos in SORNs by studying how the networks' self-organization changes their response to small perturbations. We study the effect of perturbations to the excitatory-to-excitatory weight matrix on connection strengths and on unit activities. We find that the network dynamics, characterized by an estimate of the maximum Lyapunov exponent, becomes less chaotic during its self-organization, developing into a regime where only few perturbations become amplified. We also find that due to the mixing of discrete and (quasi-)continuous variables in SORNs, small perturbations to the synaptic weights may become amplified only after a substantial delay, a phenomenon we propose to call *deferred chaos*.

Citation: Eser J, Zheng P, Triesch J (2014) Nonlinear Dynamics Analysis of a Self-Organizing Recurrent Neural Network: Chaos Waning. PLoS ONE 9(1): e86962. doi:10.1371/journal.pone.0086962

Editor: Daniel Durstewitz, Heidelberg University, Germany

Received June 7, 2013; **Accepted** December 17, 2013; **Published** January 23, 2014

Funding: This work was supported by the European Commission 7th Framework Programme (FP7/2007–2013), Challenge 2 - Cognitive Systems, Interaction, Robotics, grant agreement No. ICT-IP-231722, project "IM-CLeVeR-Intrinsically Motivated Cumulative Learning Versatile Robots" and by the LOEWE-Program Neuronal Coordination Research Focus Frankfurt (NeFF, http://www.neff-ffm.de/en/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: triesch@fias.uni-frankfurt.de

Introduction

A fundamental question in Neuroscience is how cortical circuits acquire the structure required to perform desired computations. A range of different plasticity mechanisms shape neural circuits, but their interaction at the network level remains poorly understood. Recent modeling work has shown that recurrent spiking neural networks with multiple forms of plasticity can learn interesting representations of sensory inputs [1] and reproduce experimental data on the statistics and fluctuations of synaptic connection strengths in cortex and hippocampus [2]. These self-organizing recurrent networks (SORNs) rely on an interplay of spike-timing dependent plasticity (STDP) and different homeostatic mechanisms. While these networks offer a plausible explanation for the experimentally observed approximately log-normal distribution of excitatory synaptic efficacies, their dynamics are still poorly understood. Here we try to shed light on this issue by employing tools from nonlinear dynamics analysis. Specifically, we perform a perturbation analysis, where we investigate how the evolution of a network changes in response to a small perturbation. We characterize the degree of chaos by an estimate of the maximum Lyapunov exponent and study it at different stages of network evolution.

Several previous studies have investigated the occurrence of chaos and associated irregular dynamics [3] in discrete and continuous time artificial neural networks with symmetric oscillators [4], asymmetric weight structures [5] and both weak and full connectivity [6–9]. Most studies have used sigmoidal transfer functions like the hyperbolic tangent [10–12]. The high degree of abstraction of such models often makes it hard to directly relate them to brain function, e.g. see [13]. Robust chaos is common in such systems and with more units in the network an increasing number of positive Lyapunov exponents has been observed [14,15]. The tendency for chaotic dynamics in networks with excitatory and inhibitory neurons can be reduced by Hebbian learning mechanisms as well as certain external stimuli [10,16]. Self-organization due to Hebbian learning is frequently assumed to be the responsible mechanism for the observed stabilization [10,17]. Interestingly, it has also been argued that recurrent neural networks should operate at the edge of chaos in order to maximize their computational power [18–20]. This suggests that neither highly chaotic dynamics as prevalent in random networks nor strongly regular dynamics may be desirable for the brain. Here we investigate the degree of chaos in the SORN model as studied in [2]. An interesting property of these networks is their hybrid nature that mixes discrete and (quasi-)continuous variables [21]. Our major finding is an overall reduction of chaos during its self-organization, with the network settling into a regime where only few perturbations become amplified. Due to the mixing of discrete and (quasi-)continuous variables in SORNs, they are instances of hybrid dynamical systems [21]. We also show that because of their hybrid nature, small perturbations to the synaptic weights may become amplified only after a substantial delay, a phenomenon we propose to call *deferred chaos*.

Methods

Network Model

The network model is almost identical to the one in [2]. It consists of $N^E = 200$ excitatory and $N^I = 40$ inhibitory threshold neurons linked through weighted synaptic connections. W^{EE} represents the connections between excitatory neurons. Initial connectivity is sparse with a connection probability of 0.1, but can change during network evolution. Self-connections of the excitatory neurons are forbidden. W^{EI} are the inhibitory-to-excitatory connections with connection probabilities of 0.2. W^{IE} denotes excitatory-to-inhibitory connections which are all-to-all and remain fixed at their random initial values. Connections between inhibitory neurons were left out to keep the model simple and to stay consistent with the previous work on these networks [1,2]. Unless specified otherwise, all initial weights are drawn from a uniform distribution over the interval [0,1] and then normalized such that the sum of weights projecting to a neuron is one.

The binary variables $x(t) \in \{0,1\}^{N^E}$ and $y(t) \in \{0,1\}^{N^I}$ represent the activity of the excitatory and inhibitory neurons at a discrete time t, respectively. The network activity state at time step $t+1$ is given by:

$$x_i(t+1) = \Theta\left(\sum_{j=1}^{N^E} W_{ij}^{EE}(t)x_j(t) - \sum_{k=1}^{N^I} W_{ik}^{EI}(t)y_k(t) - T_i^E(t) + \xi_E(t)\right), \quad (1)$$

$$y_i(t+1) = \Theta\left(\sum_{j=1}^{N^E} W_{ij}^{IE}x_j(t) - T_i^I + \xi_I(t)\right) . \quad (2)$$

T^E and T^I denote threshold values for the excitatory and inhibitory neurons, respectively. Initially, they are uniformly distributed in the interval $[0, T^E_{\max}]$ and $[0, T^I_{\max}]$, but not bounded during network evolution. $\Theta(\cdot)$ is the Heaviside step function. ξ_E and ξ_I represent Gaussian white noise with $\mu_\xi = 0$ and $\sigma^2_\xi = 0.05$.

The connection strength of nonzero weights between excitatory neurons is subject to the following spike-timing dependent plasticity (STDP) rule:

$$\Delta W_{ij}^{EE}(t) = \eta_{\text{STDP}}\left(x_i(t)x_j(t-1) - x_i(t-1)x_j(t)\right) . \quad (3)$$

If W_{ij}^{EE} becomes negative due to this update, the synapse is immediately eliminated.

At every time step we normalize all incoming weights to an excitatory neuron by

$$W_{ij}^{EE}(t) \leftarrow W_{ij}^{EE}(t) / \sum_j W_{ij}^{EE}(t) , \quad (4)$$

$$W_{ij}^{EI}(t) \leftarrow W_{ij}^{EI}(t) / \sum_j W_{ij}^{EI}(t) . \quad (5)$$

Furthermore, the applied intrinsic plasticity (IP) mechanism regulates the spiking thresholds in such a way that all neurons exhibit the same average firing rate H_{IP}:

$$T_i^E(t+1) = T_i^E(t) + \eta_{\text{IP}}(x_i(t) - H_{\text{IP}}) . \quad (6)$$

Thresholds of inhibitory neurons remain at their initial values.

Inhibitory weights are subject to a different STDP rule. We implement a form of inhibitory spike-timing dependent plasticity (iSTDP) for the W^{EI} connections. If an inhibitory neuron is active and the excitatory neuron receiving input from this inhibitory neuron stays silent in the following time step, the corresponding weight will be weakened by the amount η_{inhib}. If, however, the excitatory neuron is active despite receiving input from the inhibitory unit, then the weight will be strengthened by an amount $\eta_{\text{inhib}}/H_{\text{IP}}$:

$$\Delta W_{ij}^{EI}(t) = -\eta_{\text{inhib}} y_j(t-1)[1 - x_i(t)(1 + 1/H_{\text{IP}})] . \quad (7)$$

This rule aims to balance excitation and inhibition and is inspired by [22]. If an inhibitory neuron succeeds in silencing a target excitatory cell with its spike, the inhibitory influence is weakened. Conversely, if the inhibitory neuron fails to silence the excitatory cell, its inhibitory influence is strengthened.

We also introduce a structural plasticity (SP) mechanism to compensate for the synapse elimination induced by STDP. We conceive this rule as a Bernoulli experiment for each neuron. At every time step there is a small probability $p_C = 0.01$ that a neuron creates new synapses to other excitatory cells. The number of connections is drawn from a Poisson distribution with mean $\lambda_P = 2$ and their weight is set to 0.0001. The parameters used in the simulations were as follows: $\eta_{\text{STDP}} = 0.004$, $\eta_{\text{IP}} = 0.01$, $T^E_{\max} = 1$, $T^I_{\max} = 0.5$, $H_{\text{IP}} = 0.1$, $\eta_{\text{inhib}} = 0.001$.

Perturbation analysis

We apply perturbation analysis to measure the stability of the network during different stages of its evolution as observed in [2]. The network has a finite set of elements, each of which takes a discrete (unit activations) or (quasi-)continuous (weights and thresholds) value. To perform the analysis, we simulate a first network (the *original*) up to a certain time step. Then a *copy* of the network is made and a small perturbation is applied to this copy. Both original network and copy are simulated with identical noise as well as under the same influence of structural plasticity and we compare their subsequent evolution. To this end, we consider the Euclidean or Hamming distance of network weights, thresholds and activities, respectively. To construct a "baseline" Euclidean distance for the purpose of normalization due to changing statistics during network evolution, we consider the current weight matrices, threshold and input vectors of the unperturbed SORN and compute their Euclidean distance to 100 random matrices and vectors, respectively. These are generated by randomly permuting the rows/entries of the original matrix/vector. This keeps the statistics of the weights, thresholds and inputs identical. The average distance obtained in this procedure then serves as the normalization factor $1/\mathcal{N}$ when calculating the Euclidean distances. An analogous normalization for the Hamming distances is not necessary, however. As stated above, the implemented IP mechanism guarantees a constant average firing rate. Perturbations are performed by strengthening or newly generating one randomly selected entry of the W^{EE} weight matrix by an amount proportional to $\mathcal{N}$. Here, $\mathcal{N}$ corresponds to the inverse normalization factor of the W^{EE} matrix. The perturbation strength is set to $0.001 \times \mathcal{N}$.

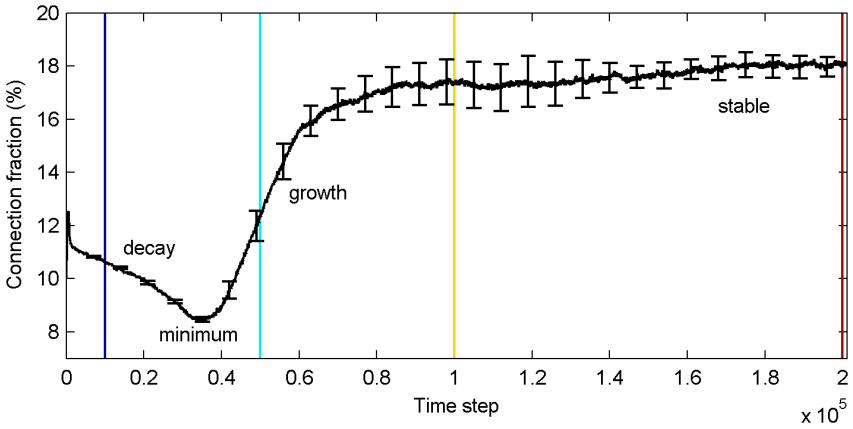

Figure 1. Average fraction of nonzero excitatory-to-excitatory connections as a function of time. The different phases of connectivity and time periods of investigation are indicated in terms of color. Errorbars represent standard errors (10 simulations) and are given every 7,000th time step.
doi:10.1371/journal.pone.0086962.g001

Note that a perturbation of the network weights will also affect the unit activities, because an increased (or decreased) synaptic weight may allow (or prevent) a neuron from reaching its firing threshold in a certain situation. In fact, since the dynamics of the weights of the network depends exclusively on the synaptic normalization and the unit activities via the STDP mechanism, a small change to one weight can only percolate to other weights after it has resulted in an effect on the discrete unit activities. Conversely, a change of unit activities will immediately affect the synaptic weights, because the firing (or silence) of a neuron may lead to the triggering (or not) of the STDP mechanism which adapts the synaptic weights. Note that the effects of perturbations to unit activities will generally be bigger, because an extra spike in a unit can lead to changes in all incoming and all outgoing connections of this neuron at the same time.

Because perturbations to the weights will only affect unit activities after a delay (at least one time step later), we define two different Lyapunov exponents. The first, immediate Lyapunov exponent Λ_i quantifies any amplifications of perturbations to the weight matrix following the perturbation as long as the neuron activities remain unaffected. The second, delayed Lyapunov exponent Λ_d quantifies amplifications of perturbations to the weight matrix after these have altered the neuron firing patterns. Specifically, let t_0 be the time of a perturbation which we observe for a total of T_t time steps and t^* be the time step right before the unit activities start differing between the perturbed and unperturbed networks ($t_0 \le t^* < t_0 + T_t$). We estimate Λ_i according to:

$$\Lambda_i = \frac{1}{T_i} \ln \frac{d_{t_0+T_i}}{d_{t_0}} , \tag{8}$$

where $T_i = t^* - t_0$ and d_t denotes the normalized Euclidean distance corresponding to the excitatory-to-excitatory weight matrix. Analogously, we estimate Λ_d as:

$$\Lambda_d = \frac{1}{T_d} \ln \frac{d_{t^*+T_d}}{d_{t^*}} . \tag{9}$$

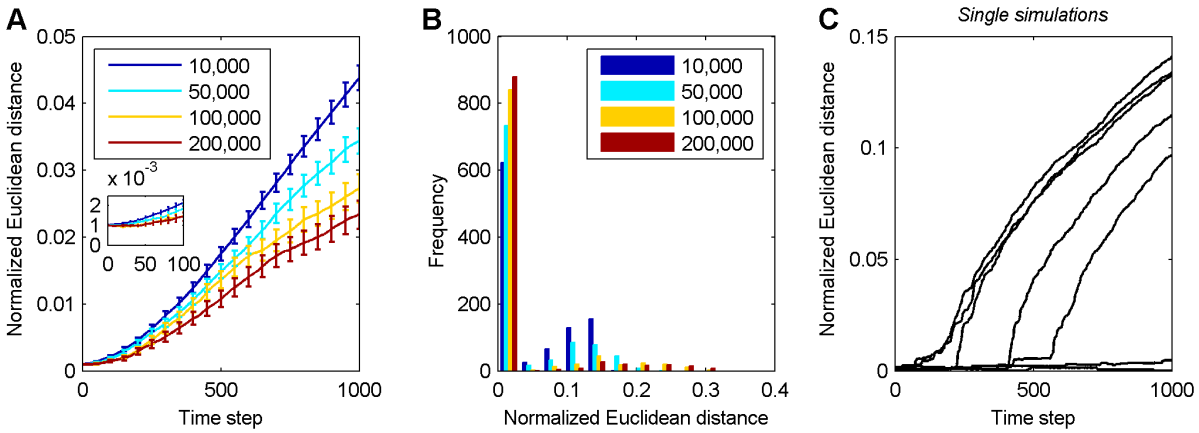

Figure 2. Long term development of normalized Euclidean distances between excitatory weight matrices of perturbed and unperturbed networks. A Averaged result over 1,000 simulations with error bars indicating standard errors of the mean. The inset shows a magnification of the initial 100 time steps. **B** Histogram of normalized Euclidean distances at 1,000th time step in A. In figures A–B, the color indicates the starting time of the perturbations as shown in the legend. **C** Examples of single simulations in the decay phase: the normalized Euclidean distance often stays at a low value for several hundred time steps but then displays an extensive amplification.
doi:10.1371/journal.pone.0086962.g002

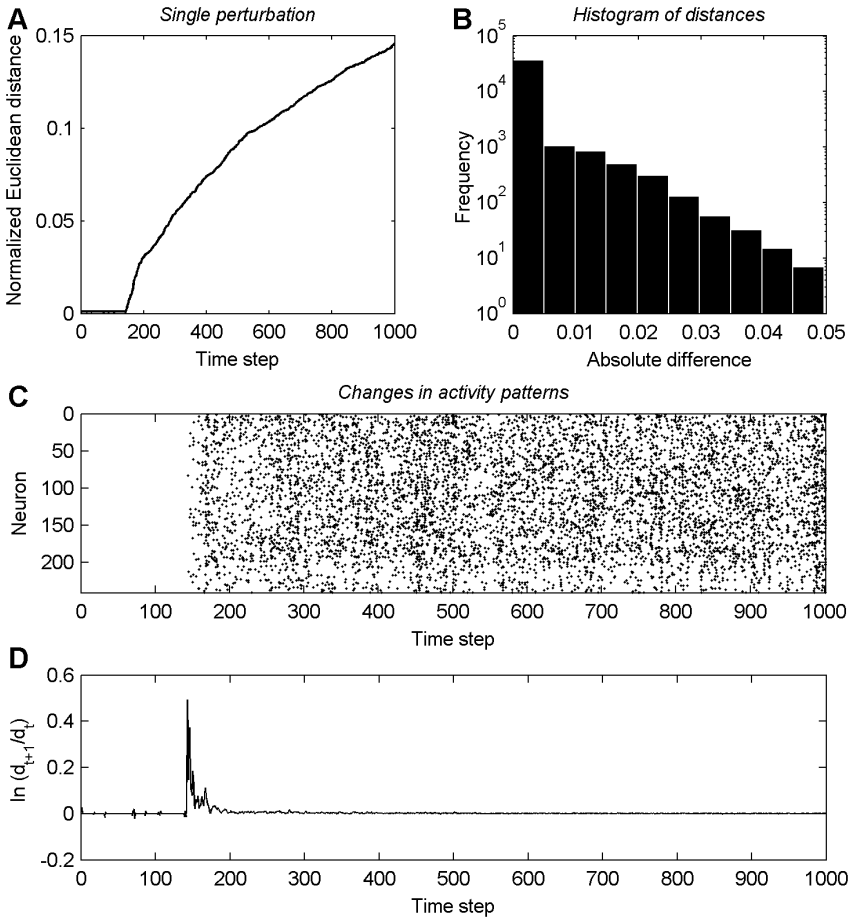

Figure 3. Example of a single amplified perturbation. A Development of the normalized Euclidean distance in a single perturbation during the decay phase. **B** Histogram of the single absolute differences of the excitatory-to-excitatory weights at time step 1,000 in A. **C** Differences in neuron activity patterns between the perturbed and the unperturbed network. **D** Time course of $\ln(d_{t+1}/d_t)$. d_t corresponds to the normalized Euclidean distance in A at time step t.
doi:10.1371/journal.pone.0086962.g003

T_d is set to 10 time steps. The combined maximum Lyapunov exponent Λ_c is computed as follows:

$$\Lambda_c = \frac{1}{T_i + T_d} \sum_{j=i,d} \Lambda_j T_j \ . \tag{10}$$

If the activities are not affected at all within the investigated time period or if they are influenced immediately after the perturbation, the combined Lyapunov exponent simply reads $\Lambda_c = \Lambda_i$ with $T_i = T_t$ and $\Lambda_c = \Lambda_d$, respectively. Else if $t^* > t_0 + T_t - T_d$, we used $T_d = t_0 + T_t - t^*$. Perturbations which completely vanish are ignored.

To characterize network dynamics in different stages of network evolution, perturbations are performed at different time points. At each stage, we consider 10 independent networks. Each network is considered at 10 different time points within 100 time steps during this phase. For each time step we perform 10 perturbations. Thus, we perform a total of $10 \times 10 \times 10 = 1{,}000$ perturbations for each stage of the network's evolution.

Results

As the network develops, its connectivity changes due to the action of the different plasticity mechanisms. Zheng et al. [2] observed that the network goes through different phases as indicated by the number of excitatory-to-excitatory connections present in the network. Fig. 1 shows the average fraction of excitatory-to-excitatory connections present in the network as a function of time. During an initial decay phase (around time step 10,000), many connections are driven to zero strength and are removed from the network. The connectivity becomes minimal between time step 30,000 and 40,000. In the subsequent growth phase, connectivity increases again. Around time step 100,000 the network has entered the stable phase where connectivity stays roughly constant.

Effect of perturbations on network weights and thresholds

We measured the long term development of the normalized Euclidean distance between the excitatory-to-excitatory weight matrix of the original and the perturbed network over 1,000 time steps after the perturbation as described in the Methods. This was done for SORNs in the three different phases of their development: starting at 10,000 time steps (decay phase), 50,000 time steps (growth phase), 100,000 and 200,000 time steps (early and late in their stable phase). The results are shown in Fig. 2 **A**. The average normalized Euclidean distance tends to increase with time. This increase is fastest for networks in the decay phase and slowest for

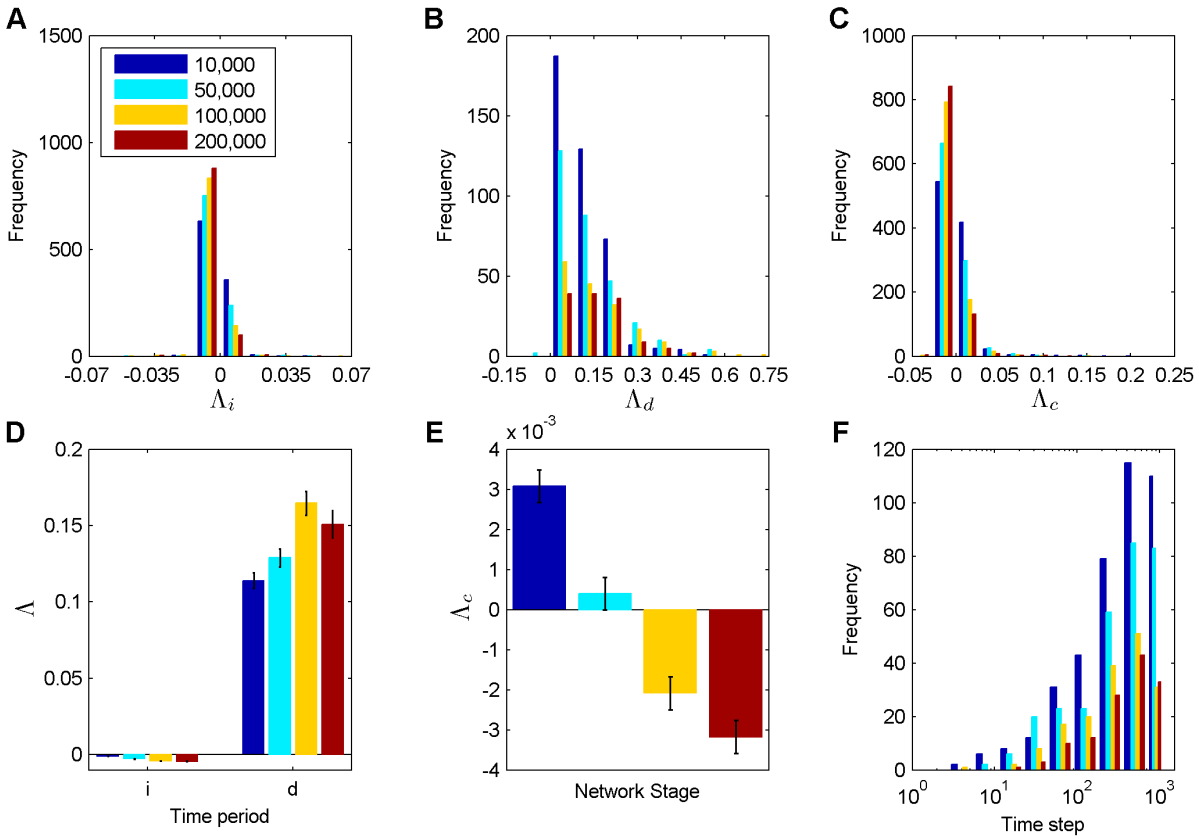

Figure 4. Deferred chaos. A Histogram of the immediate Lyapunov exponent. The legend shows the starting times of perturbations for each subfigure. **B** Histogram of the delayed Lyapunov exponent. **C** Histogram of the combined Lyapunov exponent. **D** Time course of the immediate and the delayed Lyapunov exponent. **E** Time course of the combined Lyapunov exponent. **F** Histogram of the time a weight perturbation needs to affect the neuron activities. The data correspond to Fig. 2. Errorbars in D and E indicate standard errors of the mean.
doi:10.1371/journal.pone.0086962.g004

networks in the stable phase. Fig. 2 **B** shows a histogram of normalized Euclidean distances between the original and perturbed network 1,000 time steps after the perturbation. Note that a majority of perturbations are not strongly amplified (normalized Euclidean distance stays close to zero) and only few perturbations have a strong effect on the network. During the decay phase around time step 10,000 (blue) there are many more such perturbations producing a strong effect compared to the other phases. In the first bin, the height of color bars increases monotonically as the networks go from decay to stable phase, which indicates that they become less sensitive to perturbations with time. Fig. 2 **C** gives examples of how individual perturbations during the decay phase of the network develop. The normalized Euclidean distance between the weight matrix of the original and the perturbed network tends to maintain a small value but may suddenly increase after several hundred time steps.

To illustrate this effect more clearly, we show as an example the changes of the excitatory-to-excitatory weights, the activity patterns and the time development of $\ln(d_{t+1}/d_t)$ in a single amplified perturbation from the decay phase in Fig. 3. Fig. 3 **A** displays the time course of the Euclidean distance between the weight matrices of the perturbed and unperturbed networks. Note the sharp increase around time step 150. Fig. 3 **B** shows a histogram of the single absolute differences of the excitatory-to-excitatory weights at the 1,000th time step in **A**. Interestingly, only a comparatively small fraction of weights contribute noticeably to the total normalized Euclidean distance and most of the connection strengths remain close to their original values. Data points in Fig. 3 **C** indicate situations where a neuron stays silent in the perturbed network and is active in the unperturbed network or vice versa. It is apparent that the sudden increase of the normalized Euclidean distance between both networks around time step 150 in **A** coincides both with upcoming differences in the activity patterns in **C** and an underlying time period of positive values of $\ln(d_{t+1}/d_t)$ in **D**.

The observed delayed amplification of the perturbation to the synaptic weights is due to the mixing of discrete and (quasi-)continuous variables in the network. The perturbation of a weight can only percolate to other weights after it has managed to alter one or more of the discrete unit activities. Due to the threshold function in the units' dynamics and depending on the size of the weight perturbation, this may take a substantial amount of time. We propose to call this phenomenon *deferred chaos*. Note that the increase in the Euclidean distance following changes in the unit activities in Fig. 3 **A** is not exponential. This also holds for the examples in Fig. 2 **C**. The reason for this is again that the changes to the synaptic weights are mediated by changes to the unit activities. At any time step, only a small number of weights are affected via the STDP mechanism and each of them will change its value by $\pm\eta_{\mathrm{STDP}}$. Thus the total amount of weight change per time step is limited, thereby preventing an extended exponential growth.

To quantify the *deferred chaos*, we considered the immediate, delayed and combined Lyapunov exponents introduced above.

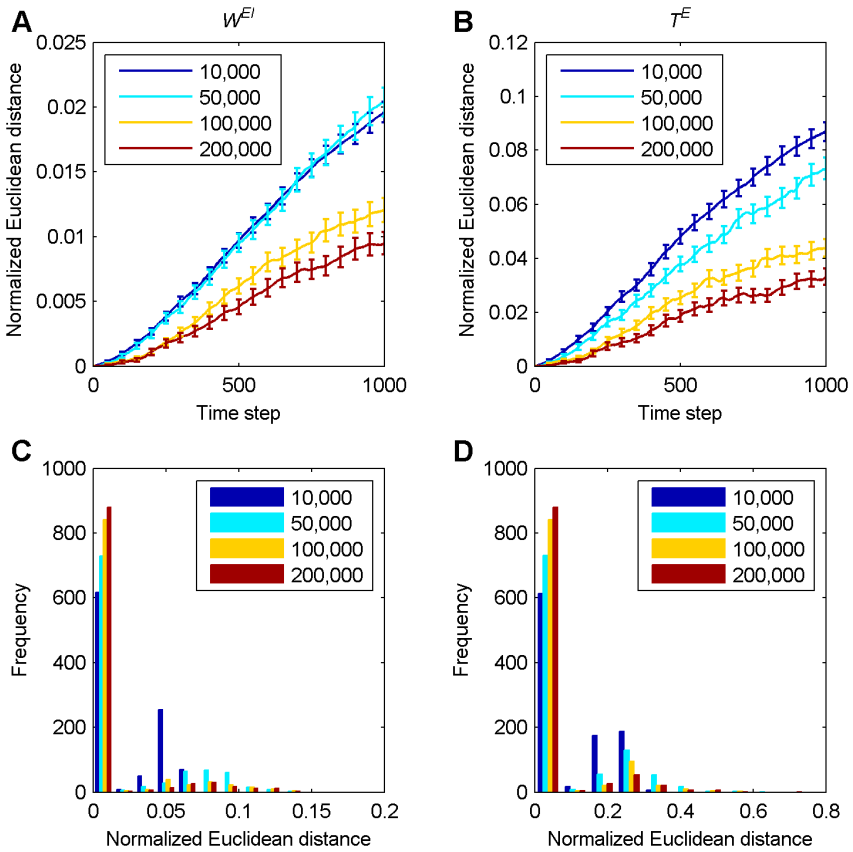

Figure 5. Long term development of normalized Euclidean distances between inhibitory-to-excitatory weight matrices and excitatory threshold vectors. The color indicates the starting time of the perturbations as shown in the legend. **A** Averaged normalized Euclidean distance between inhibitory-to-excitatory weight matrices over 1,000 simulations for 1,000 time steps. **B** Averaged normalized Euclidean distance between excitatory threshold vectors over 1,000 simulations for 1,000 time steps. Errorbars indicate standard errors of the mean in A and B. **C** Histogram of normalized Euclidean distances at 1,000th time step in A. **D** Histogram of normalized Euclidean distances at 1,000th time step in B.
doi:10.1371/journal.pone.0086962.g005

Fig. 4 **A**–**C** show histograms of Λ_i, Λ_d and Λ_c at different stages of network evolution. The majority of perturbations produce a negative value of Λ_i across all phases. We also find more perturbations that have a slightly positive value during the decay phase (around time step 10,000) than in any other phase. Especially, the stable phase (around time steps 100,000 and 200,000) displays a huge amount of slightly negative values. Fig. 4 **B** shows histograms of estimated Λ_d values for the different phases of network evolution. Almost all estimates are positive, indicating chaotic behavior right after a perturbation has affected the unit activities. Note that only a fraction of perturbations ever affect the unit activities, however. This fraction is highest in the decay phase around time step 10,000 and becomes smaller with network evolution (10,000: 40.6%, 50,000: 30.1%, 100,000: 16.9%, 200,000: 13.0%). Also note that since the growth of the Euclidean distance of the weight matrix after "take-off" is generally not exponential as explained above (compare Fig. 3 **A**), the estimate of Λ_d depends on the time interval T_d over which it is estimated. Results are shown for $T_d = 10$ time steps (compare eq. (9)). Longer time intervals will lead to a smaller estimate of Λ_d. Fig. 4 **C** shows histograms of Λ_c values for the different phases of network evolution, which are estimated by combining Λ_i and Λ_d according to (10). The number of negative Λ_c estimates is smallest for networks in the decay phase and highest for networks in the stable phase.

Fig. 4 **D**, **E** show the average Λ_i, Λ_d, and Λ_c for the four different stages of network evolution. Λ_i, Λ_d and Λ_c are significantly different from zero across all phases (except Λ_c during the growth phase: $p = 0.3230$, else: $p < 10^{-7}$; t-test). On average, Λ_i (Λ_d) is strictly negative (positive) during the whole network evolution. Λ_d significantly increases from growth to stable phase. However, we also find a significant decrease of Λ_i and Λ_c from decay to stable phase ($p < 10^{-6}$ for Λ_i, Λ_d and Λ_c; ANOVA with multiple comparisons and post-hoc t-test), indicating chaos waning. Specifically, Λ_c starts at a positive value during the decay phase, crosses the critical line and settles down to negative values when entering the stable phase. Finally, we find a significant increase according to $\Lambda_i < \Lambda_c < \Lambda_d$ averaged across all stages of network evolution ($p < 10^{-15}$; ANOVA with multiple comparisons and post-hoc t-test). Only 16 out of a total of 4×1000 perturbations vanished completely and were excluded, which should not strongly bias the results.

We also quantified the distribution of time periods of how long it takes a weight perturbation to influence the neuron activities. Fig. 4 **F** shows histograms of such "take-off" times depending on the stage of network evolution. Across all network stages the perturbations typically need several hundred time steps to alter activities for the first time. Note that the relatively steady increase of the normalized Euclidean distance in Fig. 2 **A** results from averaging many curves like the ones in Fig. 2 **C** with different "take-off" times.

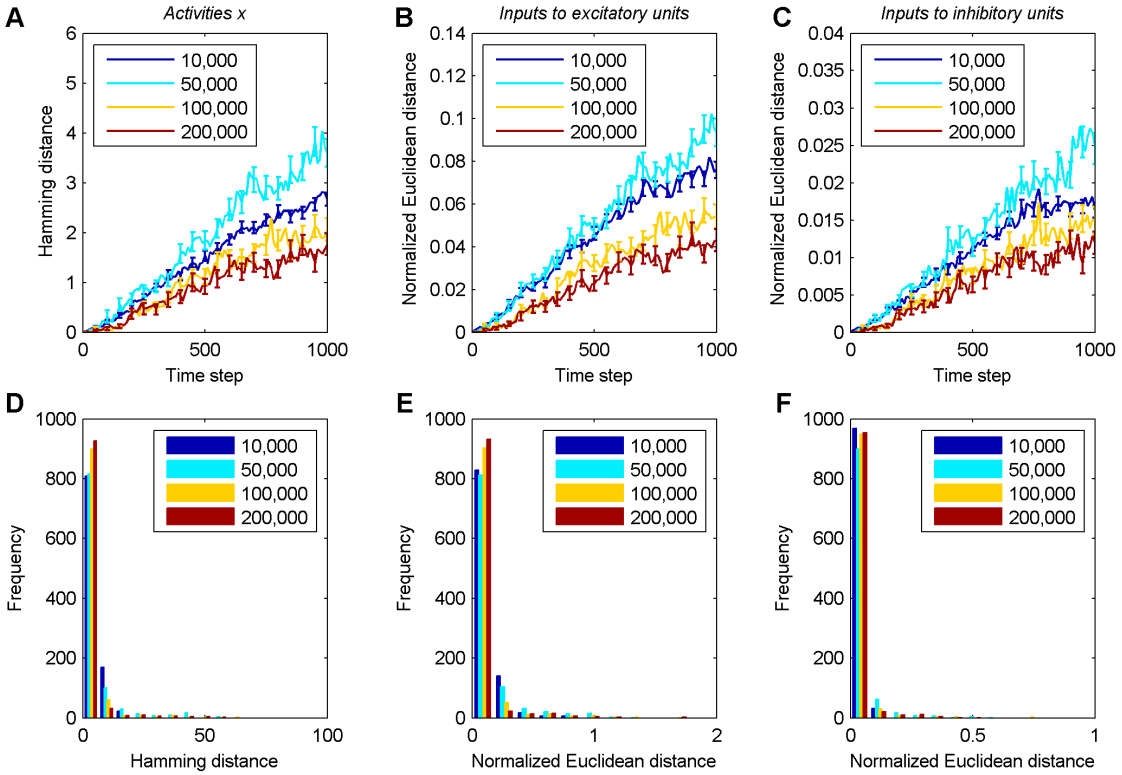

Figure 6. Perturbation analysis of neuron activities. The legend shows the starting times of perturbations. **A** Averaged Hamming distance between neuron activities of the perturbed and unperturbed networks over 1,000 simulations for 1,000 time steps. Only every tenth time step is plotted. Error bars indicate standard errors of the mean in A–C and are given every 50th time step. **B** Averaged normalized Euclidean distance of excitatory neuron inputs over 1,000 simulations for 1,000 time steps. **C** Averaged normalized Euclidean distance of inhibitory neuron inputs over 1,000 simulations for 1,000 time steps. **D**–**F** Histograms of Hamming distances and normalized Euclidean distances at 1,000th time step in A–C, respectively.
doi:10.1371/journal.pone.0086962.g006

Fig. 5 shows how a perturbation to the excitatory-to-excitatory weights affects inhibitory-to-excitatory connections and excitatory thresholds. Again, perturbations are amplified least when the network has reached the stable phase around time step 100,000 and 200,000. In Fig. 5 **A**, there is no significant difference between the decay phase (around time step 10,000) and the growth phase (around time step 50,000), which is in contrast to the development of the excitatory-to-excitatory weights as shown in Fig. 2 **A** and the thresholds as shown in Fig. 5 **B**. However, part **C** and **D** of Fig. 5 exhibit the same stabilizing phenomenon as Fig. 2 **B**: the fraction of perturbations that are not strongly amplified increases monotonically across the different stages of network evolution.

Effect of perturbations on the unit activities

In Fig. 3 **C** we already saw how a single perturbation to the synaptic weights can influence the unit activities. To study this effect more systematically, we consider the evolution of the Hamming distance between the activities of the excitatory units in the original and the perturbed network. Fig. 6 **A** shows that the average long term development of the Hamming distance increases with time during all phases. Consistent with the results from the previous section, the amplification of an initial perturbation tends to become smaller as the network develops. In the stable phase, perturbations grow most slowly. A similar behavior is observed for the non-discrete inputs to the excitatory and inhibitory units. Specifically, we consider the normalized Euclidean distance between the arguments of the Heaviside step functions in the update equations (1) and (2) for the excitatory (Fig. 6 **B**) and inhibitory units (Fig. 6 **C**). Note that the effects of the perturbation are generally small. A mean Hamming distance of 4.5 after 1,000 time steps (compare Fig. 6 **A**) means that on average only four or five units out of 200 have a different activity state. Often the perturbation is immediately "forgotten" in the next time step and only infrequently it is amplified up to a limiting mean value of 36 corresponding to two independent networks with sparse activity of 10% active units per time step. Nevertheless, we find once again in two out of three cases the smallest amount of strongly amplified simulations in the stable phase, which is evidence for the network's stabilization (compare Fig. 6 **D**–**F**).

We speculated that the effect of a perturbation to a synaptic weight may depend on the outdegree of the receiving neuron. A unit with a high outdegree connects to many other units in the network. Therefore, altering one of its incoming weights may affect a large part of the network. This raises the question if the normalized Euclidean distance is correlated to the outdegree of the neuron that receives the perturbed weight. This is not the case, however. Fig. 7 shows a scatter plot of the perturbed neuron's outdegree vs. the normalized Euclidean distance between excitatory weight matrices of original and perturbed network after 1,000 time steps for 1,000 perturbations in each of the three phases of network development. Coefficients of correlation ρ are small and not significant at the 5% level in any of the four stages (10,000 time steps: $\rho = -0.0106$, $p = 0.7378$; 50,000 time steps: $\rho = 0.0050$, $p = 0.8745$; 100,000 time steps: $\rho = -0.0613$, $p = 0.0526$; 200,000 time steps: $\rho = 0.0582$, $p = 0.0658$; two-tailed t-test).

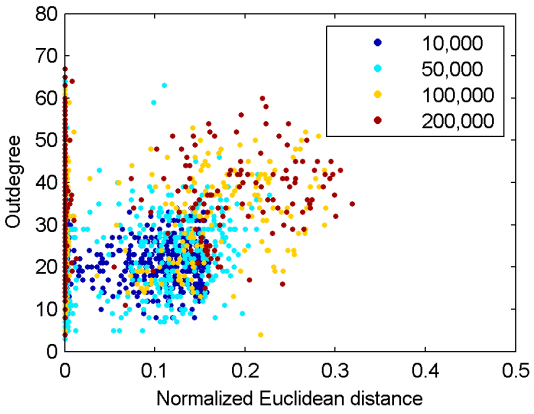

Figure 7. Relationship of outdegree vs. normalized Euclidean distance at 1,000th time step in Fig. 2 A. Scatter plot of the perturbed neurons' outdegrees and the corresponding normalized Euclidean distances between the excitatory weight matrices. Coefficients of correlation between the outdegree and the normalized Euclidean distance are not significantly different from zero.
doi:10.1371/journal.pone.0086962.g007

Discussion

The potential role of chaos in the brain has long been a topic of debate. On the one hand, recent work suggests that chaotic dynamics could play a central role in approximating stochastic inference schemes, e.g., [23]. On the other hand, recent theoretical considerations and experimental data support the view that the brain operates in a regime close to criticality that is neither highly chaotic nor strongly regular [18–20,24]. A recent analysis by Priesemann et al. indicates that the human brain may in fact favor a slightly subcritical operating regime across different vigilance states [25]. This raises the question how the brain tunes its dynamics towards such a slightly subcritical regime.

Here we have investigated the dynamics of the recently proposed self-organizing recurrent neural network model (SORN) [1,2]. This model has been shown to learn effective representations of dynamic inputs and to reproduce experimental data on the statistics and fluctuations of synaptic connection strengths in cortex and hippocampus. Here we have studied the network's self-organization without any structured input but under the influence of weak noise. To characterize the network's dynamics during different phases of its self-organization, we performed a perturbation analysis and quantified the sensitivity of the network to small changes to its synaptic weights. Our major findings are that only a fraction of perturbations becomes amplified within a limited time window, and this fraction decreases across the different phases of network self-organization. This is consistent with findings in other recurrent network architectures with Hebbian-like plasticity mechanisms showing a development towards more stable dynamics [10,14,26]. Interestingly, most perturbations to the synaptic weights remained at a "critical" level or even disappeared again. We showed that amplification of a perturbation depends on whether it affects the neural activity or not. Perturbations that are amplified produce drastic changes in the network's firing patterns compared to the cases where neuron activity in perturbed and unperturbed network stays identical. Importantly, we found that such changes in the network's activity patterns could occur after substantial delays, a phenomenon we refer to as *deferred chaos*. This effect is a consequence of the hybrid nature of our network in that it mixes discrete (neuron activities) and (quasi-)continuous (weights and thresholds) variables [21]. Since changes to the synaptic weights are driven by the neurons' activity patterns, a change to a synaptic connection has to first affect the activitiy of the recipient neuron before it can percolate to other weights in the network. Depending on the size of the initial perturbation this can take a very long time across all different phases of network self-organization. It is an open question whether there are interesting engineering applications for such *deferred chaos*.

Overall, our results are broadly consistent with earlier findings that similar networks may develop dynamics that are close to the critical regime [26]. Further analysis is needed to understand how the network's dynamics is influenced by structured inputs.

Acknowledgments

The authors thank the anonymous reviewers for helpful suggestions.

Author Contributions

Conceived and designed the experiments: JE JT PZ. Performed the experiments: JE. Analyzed the data: JE. Wrote the paper: JE JT PZ.

References

1. Lazar A, Pipa G, Triesch J (2009) SORN: a self-organizing recurrent neural network. Frontiers in Computational Neuroscience 3: 23.
2. Zheng P, Dimitrakakis C, Triesch J (2013) Network Self-Organization Explains the Statistics and Dynamics of Synaptic Connection Strengths in Cortex. PLoS Computational Biology 9: e1002848.
3. Jahnke S, Memmesheimer RM, Timme M (2008) Stable Irregular Dynamics in Complex Neural Networks. Physical Review Letters 100: 048102.
4. Bick C, Timme M, Paulikat D, Rathlev D, Ashwin P (2011) Chaos in Symmetric Phase Oscillator Networks. Physical Review Letters 107: 244101.
5. Sompolinsky H, Crisanti A, Sommers HJ (1988) Chaos in Random Neural Networks. Physical Review Letters 61: 259–262.
6. Albers DJ, Sprott JC, Dechert WD (1998) Routes to chaos in neural networks with random weights. International Journal of Bifurcation and Chaos 8: 1463–1478.
7. Siri B, Quoy M, Delord B, Cessac B, Berry H (2007) Effects of Hebbian learning on the dynamics and structure of random networks with inhibitory and excitatory neurons. Journal of Physiology - Paris 101: 136–148.
8. Pasemann F (2002) Complex dynamics and the structure of small neural networks. Network: Computational Neural Systems 13: 195–216.
9. Potapov A, Ali MK (2000) Robust chaos in neural networks. Physics Letters A 277: 310–322.
10. Dauce E, Quoy M, Cessac B, Doyon B, Samuelides M (1998) Self-organization and dynamics reduction in recurrent networks: stimulus presentation and learning. Neural Networks 11: 521–533.
11. Chen L, Aihara K (1997) Chaos and asymptotical stability in discrete-time neural networks. Physica D 104: 286–325.
12. Cessac B, Doyon B, Quoy M, Samuelides M (1994) Mean-field equations, bifurcation map and route to chaos in discrete time neural networks. Physica D 74: 24–44.
13. Doyon B, Cessac B, Quoy M, Samuelides M (1994) On bifurcations and chaos in random neural networks. Acta Biotheoretica 42: 215–225.
14. Sprott JC (2008) Chaotic dynamics on large networks. Chaos 18: 023135.
15. Albers DJ, Sprott JC, Crutchfield JP (2006) Persistent chaos in high dimensions. Physical Review E 74: 057201.
16. Rajan K, Abbott LF, Sompolinsky H (2010) Stimulus-dependent suppression of chaos in recurrent neural networks. Physical Review E 82: 011903.
17. Sprott JC (2008) Simple models of complex chaotic systems. American Journal of Physics 76: 474–480.
18. Shew WL, Plenz D (2013) The Functional Benefits of Criticality in the Cortex. The Neuroscientist 19: 88–100.
19. Bertschinger N, Natschläger T (2004) Real-Time Computation at the Edge of Chaos in Recurrent Neural Networks. Neural Computation 16: 1413–1436.
20. Legenstein R, Maass W (2007) Edge of chaos and prediction of computational performance for neural circuit models. Neural Networks 20: 323–334.
21. Aihara K, Suzuki H (2010) Theory of hybrid dynamical systems and its applications to biological and medical systems. Philosophical Transactions of the Royal Society A 368: 4893–4914.
22. Vogels TP, Abbott LF (2009) Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. Nature Neuroscience 12: 483–491.
23. Suzuki H, Imura J, Horio Y, Aihara K (2013) Chaotic Boltzmann machines. Scientific Reports 3: 1610.

24. Kitzbichler MG, Smith ML, Christensen SR, Bullmore E (2013) Broadband Criticality of Human Brain Network Synchronization. PLoS Computational Biology 5: e1000314.
25. Priesemann V, Valderrama M, Wibral M, Le Van Quyen M (2013) Neuronal Avalanches Differ from Wakefulness to Deep Sleep-Evidence from Intracranial Depth Recordings in Humans. PLoS Computational Biology 9: e1002985.
26. Lazar A, Pipa G, Triesch J (2007) Fading memory and time series prediction in recurrent networks with different forms of plasticity. Neural Networks 20: 312–322.