Supplementary Information

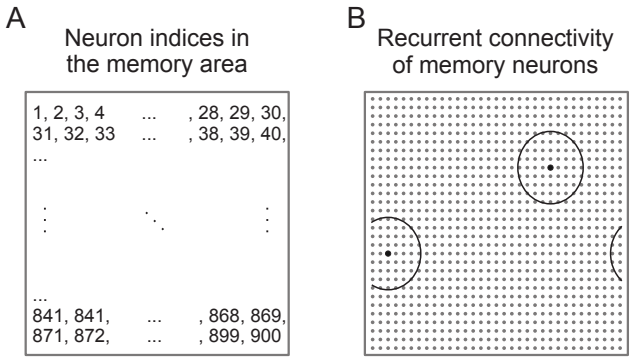
Auth et al.

February 6, 2018

SI Section A: Network Simulation Details

Memory Area Architecture

For simplicity, we consider locally connected neurons in the memory area. For this, the neurons are arranged on a 30x30-grid with indexes running from left-top to right-bottom (Figure S1 A). Each neuron is connected to its neighbors, if the position of the neighboring neuron is within a circle of radius equals 4 (see two examples in Figure S1 B). Periodic boundary conditions are introduced to avoid boundary and finite-size effects.



S 1. Arrangement of neurons and recurrent connectivity: (A) The indexes of the neurons in the memory area are running from the top-left to the the bottom-right. (B) Exemplary illustration of the connectivity of two neurons (bigger black dots) in the memory area.

Measures

Average shortest path length

The *average shortest path length* (ASPL, Figure 1 D) is considered here as a measure to analyze the spatial distribution of activation within the memory area. A high ASPL between neurons indicates that these neurons are spatially broadly distributed across the memory area. By contrast, a low ASPL indicates that the neurons are clustered. In particular, as strongly activated neurons are supposed to become part of a memory representation, we focus on the distribution of highly activated neurons. For this, for each trial, we identified the 10% of neurons with the highest activity level (index set P) and calculated the shortest path length (SPL; see 49 in main text citations); using the *networkX* package for *Python*) between them and averaged over all those paths (denoted by $\langle\cdot\rangle$):

$$\mathrm{ASPL} = \langle \mathrm{SPL_{i,j}} \rangle_{i,j \in P, \mathrm{i} \neq \mathrm{j}}. \tag{1}$$

The *average outgoing recurrent synaptic weight* (Figure 1 E) is a measure of the interconnection within a neuronal subpopulation in the memory area (index set Q). We therefore averaged the synaptic weight over all the connections among neurons within the sup-population:

$$\bar{w}^{\mathrm{rec}} = \langle w^{\mathrm{rec}}_{\mathrm{i,j}} \rangle_{i,j \in Q, \mathrm{i} \neq \mathrm{j}}. \tag{2}$$

The *average incoming feed-forward synaptic weight* (Figure 1 F,G) is the average synaptic weight of connections between a subpopulation in the memory area (index set Q) and a specific stimulus pattern in the input area (index set H):

$$\bar{w}^{\mathrm{ff}} = \langle w^{\mathrm{ff}}_{\mathrm{i,k}} \rangle_{i \in Q, k \in H}. \tag{3}$$

Note that the synaptic weights determined by the interplay between Hebbian synaptic plasticity and synaptic scaling are normalized according to the weight values $\hat{w}^{\mathrm{ff}}$, $\hat{w}^{\mathrm{rec}}$ resulting from maximum activation levels. Thus, we calculated the fixed points of equations 5 and 6 in the main text given that $F_i = F_j = \alpha$ and $I_k = 130$:

$$\hat{w}^{\mathrm{ff}} = \sqrt{\frac{\kappa^{\mathrm{ff}} \alpha \cdot 130}{\alpha - F^{\mathrm{T}}}}, \tag{4}$$

$$\hat{w}^{\mathrm{rec}} = \sqrt{\frac{\kappa^{\mathrm{rec}} \alpha^2}{\alpha - F^{\mathrm{T}}}}. \tag{5}$$

Response disparity dependent on stimulus similarity

To analyze the response disparity, stimulus A is presented 10 times for 5 sec with 1 sec pause in between to form a single CA. After that, plasticity is shut off and the system is presented variations of stimulus A with increasing *stimulus disparity* (Figure 1 H). Stimulus disparity measures the relative amount of non-overlap between two stimulus patterns, in this case stimulus A and its variation (in the following called stimulus A'). Both stimuli are of identical size $N^{\mathrm{S}} = 0.5 \cdot N^{\mathrm{I}}$, so that the stimulus disparity is calculated as follows:

$$\textit{stimulus disparity}\,(A,\, A') \,=\, 1 - \frac{1}{N^{\mathrm{S}}} \cdot \sum_{\mathrm{k}}^{N^{\mathrm{I}}} S_{\mathrm{k}}(A) \cdot S_{\mathrm{k}}(A') \tag{6}$$

with binary stimulus patterns for a given stimulus $X \in A, A'$:

$$S_{\mathrm{k}}(X) = \begin{cases} 1, & \text{if } I_{\mathrm{k}}(X) \,=\, 130, \\ 0, & \text{if } I_{\mathrm{k}}(X) \,=\, 0. \end{cases} \tag{7}$$

Thus, a stimulus disparity equal zero describes two identical stimuli, whereas a disparity equal one indicates two non-overlapping stimulus patterns. The input area size of $N^{\mathrm{I}} = 36$ allows for 18 steps in variation of $5.\bar{5}\%$ each. At the end of each presentation we compare the resulting response in the memory area with the one at the end of the learning phase (i.e. the response to the original stimulus A). The *response vector overlap* (RVO; Figure 1 H) describes the similarity between the response patterns in the memory area due to the presentation of stimuli A and A':

$$RVO(A,\, A') \,=\, \sum_{\mathrm{i}}^{\mathrm{N^M}} R_{\mathrm{i}}(A) \cdot R_{\mathrm{i}}(A') \tag{8}$$

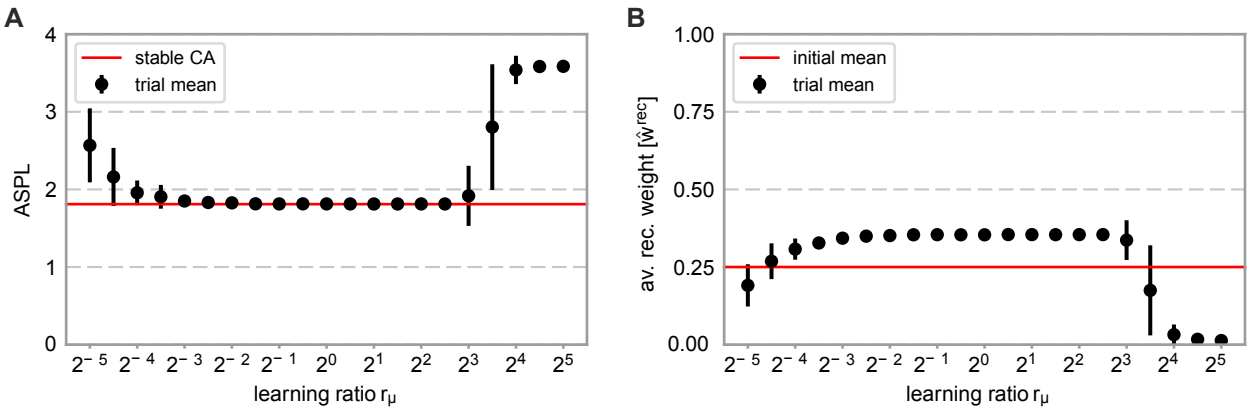
with binary response of neuron i to a given stimulus $X \in A, A'$:

$$R_{\mathrm{i}}(X) = \begin{cases} 1, & \text{if } F_{\mathrm{i}}(X) \geqslant 0.5 \cdot \alpha, \\ 0, & \text{else.} \end{cases} \tag{9}$$

Parameter Dependency of CA Formation

The formation of a cell assembly in the neural circuit depends on the synaptic weight dynamics at two sites: 1) the feed-forward synaptic connections from the input area to the memory area and 2) the recurrent connections among neurons of the memory area. In the main text, both plasticities (feed-forward as well as recurrent) act on the same time scale $\mu^{\mathrm{ff}} = \mu^{\mathrm{rec}} = \mu$ (compare equations 5 and 6). Here, we show that the formation of a CA is robust against parameter changes in a wide regime. In general, we consider a CA to be a strongly interconnected subpopulation of neurons with highly correlated activities, i.e. simultaneously high firing rates. Therefore, we show the resulting ASPL between the 10% strongest active neurons in the memory area (Figure S2 A) and the average recurrent synaptic weight the subpopulation receives (Figure S2 B) for variations of the time scale ratio of the feed-forward and recurrent weight adaptation $r_\mu = \frac{\mu^{\mathrm{rec}}}{\mu^{\mathrm{ff}}}$. The learning protocol is equivalent to the first learning phase used for Figure 1 D in the main text. The ASPL and the average weight are evaluated after the last stimulus presentation.

A low ASPL (≈ 1.8; Figure S2 A, red line) and a significant increase of the recurrent synaptic weights above the initial mean value (0.25; Figure S2 B, red) indicate a proper formation of a CA. These two requirements are fulfilled for a wide range of the learning ratio ($r_\mu = 2^{-4}$ to $r_\mu = 2^{3.5}$).

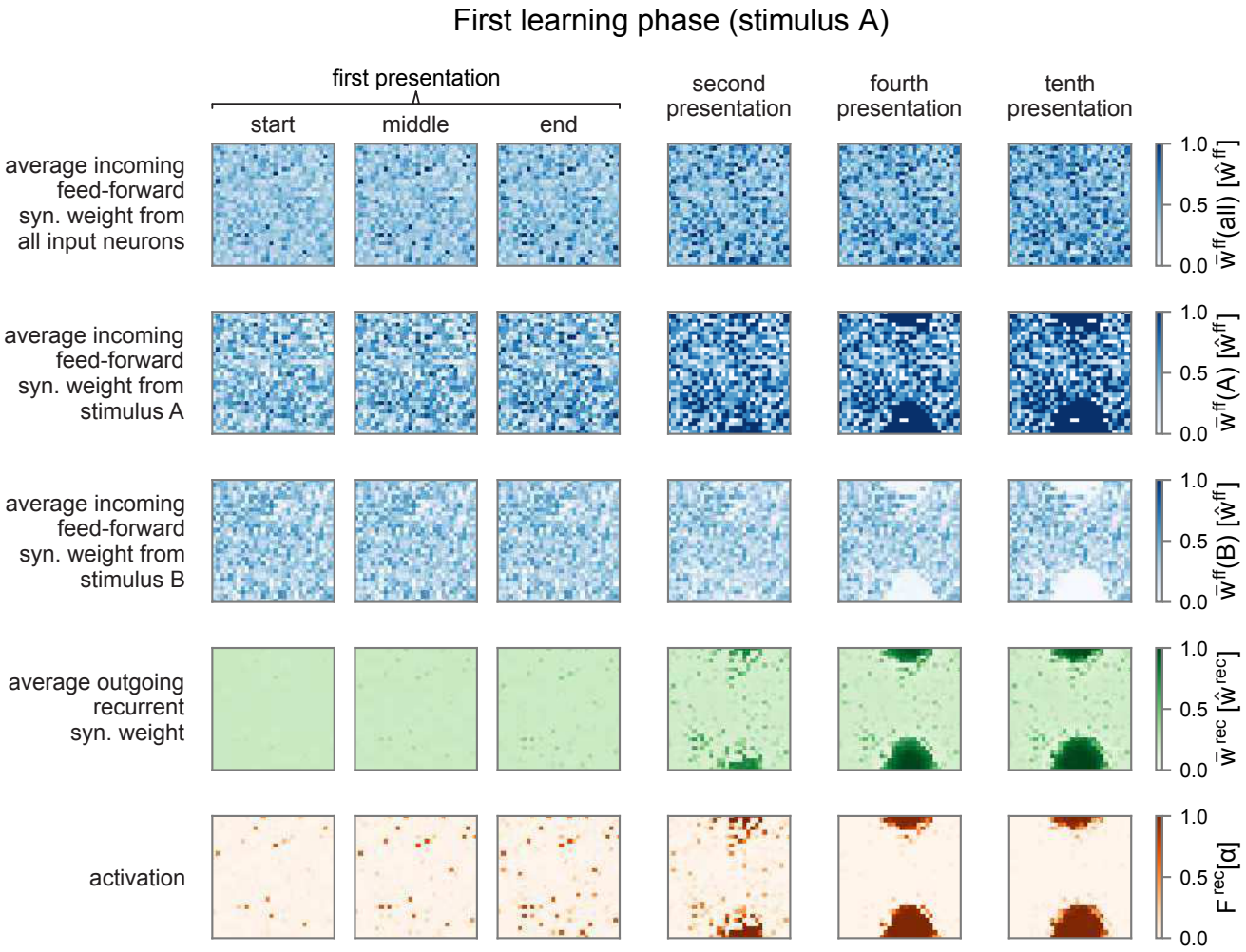


S 2. The parameter-dependency of different plasticity time scales on the CA formation. (A) The average shortest path length (ASPL) between the 10% of strongest active neurons in the memory area versus the time scale ratio r_μ. (B) The average synaptic weight of all recurrent synapses this active subpopulation receives versus r_μ. Data points are means±standard deviation over 100 trials. For more details see text.

System Development

In this section, we show the development of the neural circuit in more detail. Figures S3 and S4 capture key values of a learning trial for each neuron in the memory area at different points in time. For illustrative reasons the learning protocol is split into two parts: Figure S3 shows dynamics during the the first learning phase (stimulus A), while Figure S4 shows them during the second learning phase (stimulus B).

Each square represents the $30x30$-grid of neurons in the memory area as indicated in Figure S1 A. Thus, each point shows the average values of a neuron's feed-forward synaptic weights (first row: from all input neurons to corresponding neuron; second row: from stimulus A to corresponding neuron; third row: from stimulus B to corresponding neuron), outgoing recurrent synaptic weights (fourth row) and its neuronal activation (fifth row).

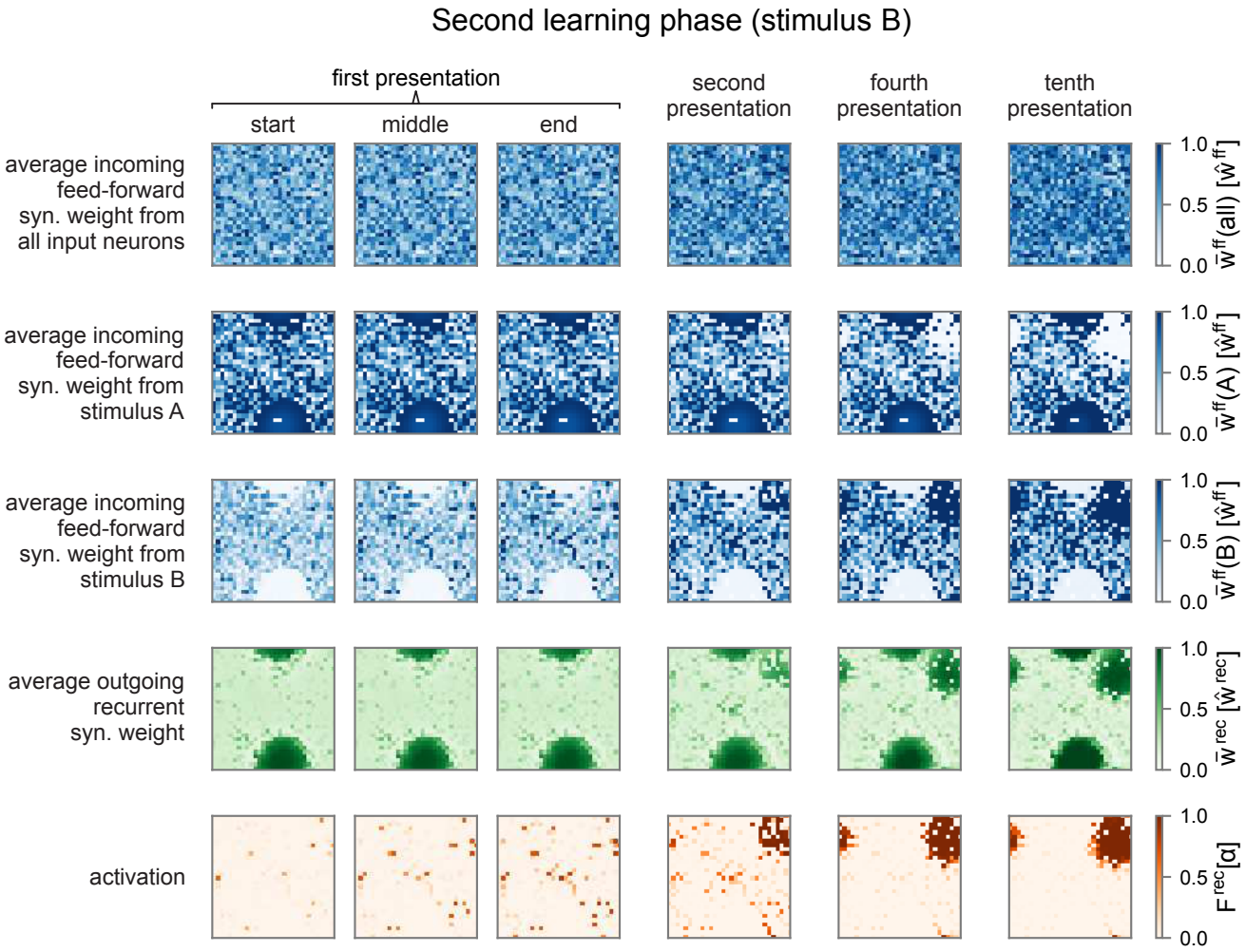


S 3. System state of the memory area at different points in time during the first learning phase. For details see text.

First learning phase: the system is initialized with zero activation in the memory area, recurrent weights are at $0.25 \cdot \hat{w}^{\text{rec}}$ and feed-forward weights are drawn from a uniform distribution between 0 and

$0.7 \cdot \hat{w}^{\text{ff}}$. Presenting stimulus A to the memory area initially activates an increasing number of broadly distributed neurons in the memory area (red dots in the fifth row of Figure S3; first presentation). The ongoing activity increases recurrent synaptic weights (green; fourth row) and also feed-forward weights from stimulus A (blue; second row), which, in turn, increases the activity. This positive feedback loop between synaptic weights and neuronal activity leads to the emergence of clustered activity (second presentation). This cluster spreads until inhibition limits its growth and, furthermore, suppresses sparse activation in the periphery (fourth to tenth presentation). The recurrent weights among CA neurons are increased until the equilibrium between Hebbian synaptic plasticity and synaptic scaling is reached. Note that, if averaged over all of the input area's connections (blue; top row), the synaptic weight changes of the feed-forward connections to the CA are not significant. By contrast, the synaptic weights only from the stimulus-A-neurons to the emerging CA-neurons are strengthened (blue spot; second row). Furthermore, synaptic weights from stimulus-B-neurons to the CA-neurons are decreased (white spot; third row).

Second learning phase: the first learning phase yields the encoding of a highly interconnected subpopulation of neurons in the memory area. However, due to the interplay between Hebbian synaptic plasticity and synaptic scaling this CA cannot be activated by the second stimulus B. Instead, the process of initially scattered activation (red dots in the fifth row of Figure S4; first presentation) and the following clustering process (fourth to tenth presentation), as described above, are repeated yielding the formation of a second CA representing stimulus B.

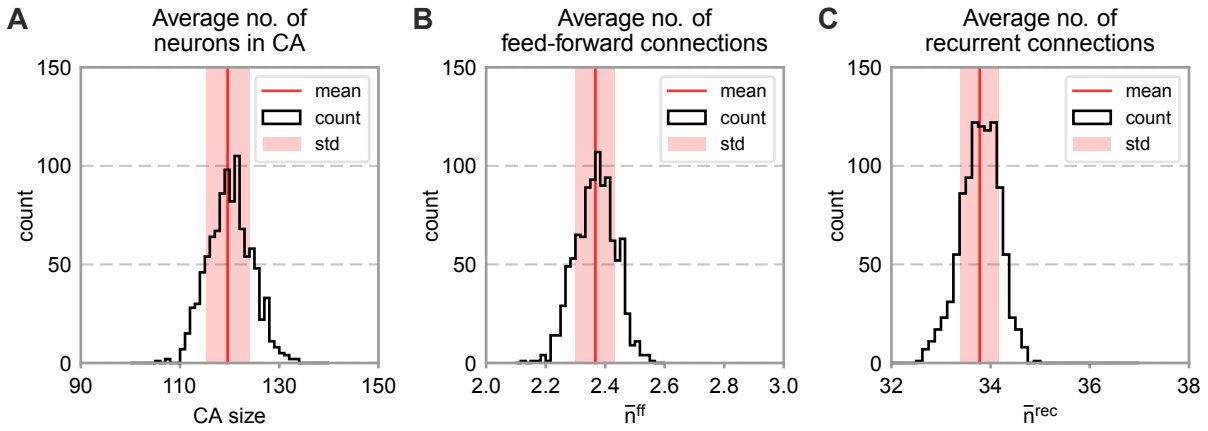


S 4. System state of the memory area at different points in time during the second learning phase. For details see text.

Cell Assembly Topology

Some of the parameters of the population model are extracted from the full network simulations. Average values are given as *mean*±*standard deviation*.

Figure S5 A shows the distribution of numbers of neurons within the first CA with average $\bar{N}^{\mathrm{CA}} = 120 \pm 4$. Furthermore, the average number of feed-forward connections per CA-neuron $\bar{n}^{\mathrm{ff}}$ is calculated. Figure S5 B shows the distribution for $\bar{n}^{\mathrm{ff}}$ from stimuli A to the corresponding CA. Stimulus A consists of half the input population being active. Thus, with an exact number of four randomly assigned feed-forward connection per memory neuron, every stimulus-A-neuron should on average project with two feed-forward connections on each neuron in the memory area. Due to the random connectivity, however, stimulus-A-neurons project to the corresponding CA-neurons with $\bar{n}^{\mathrm{ff}} = 2.37 \pm 0.07$. We also determined the average number of recurrent connections $\bar{n}^{\mathrm{rec}}$ each neuron within the first CA receives from other CA-neurons (Figure S5 C; $\bar{n}^{\mathrm{rec}} = 33.8 \pm 0.4$).



S 5. Different distributions of CA properties for 1000 trials. (A) Distribution of number of neurons in a CA, (B) of the average number of incoming feed-forward connections per CA neuron from the corresponding stimulus, (C) and of the average number of incoming recurrent connections per CA-neuron.

SI Section B: Population Model Details

Nullclines and Equilibria

The equilibrium values $\bar{w}_i^{\mathrm{rec},*}$ and $\bar{w}_i^{\mathrm{ff},*}$ of the recurrent and feed-forward weights can be obtained as a function of the equilibrium values of the population activities $\bar{F}_i^*$ from equations 10 and 11 of the main text:

$$\bar{w}_i^{\mathrm{rec},*}(\bar{F}_i^*) = \sqrt{\frac{\kappa^{\mathrm{rec}}(\bar{F}_i^*)^2}{\bar{F}_i^* - F^{\mathrm{T}}}}, \tag{10}$$

$$\bar{w}_{ik}^{\mathrm{ff},*}(\bar{F}_i^*) = \sqrt{\frac{\kappa^{\mathrm{ff}}\bar{F}_i^*\bar{I}_k}{\bar{F}_i^* - F^{\mathrm{T}}}}. \tag{11}$$

The equilibrium value $\bar{u}_{\mathrm{inh}}^*$ of the membrane potential of the inhibitory population can be formulated as a function of $\bar{F}_1^*$ and $\bar{F}_2^*$ based on equation 8 in the main text:

$$\bar{u}_{\mathrm{inh}}^*(\bar{F}_1^*, \bar{F}_2^*) = R_{\mathrm{inh}}\tau_{\mathrm{inh}}(w_{\mathrm{inh},1}N_1\bar{F}_1^* + w_{\mathrm{inh},2}N_2\bar{F}_2^*). \tag{12}$$

By inserting equations 10, 11 and 12 into equation 7 in the main text and using $\bar{F}_i^* = F(\bar{u}_i^*)$, we obtain a system of the two population nullclines that only depends on the equilibrium values $\bar{u}_1^*$ and $\bar{u}_2^*$ ($i \in \{1, 2\}$):

$$0 = -\frac{\bar{u}_i^*}{\tau} + R\left(\bar{n}_i^{\mathrm{rec}}\bar{w}_i^{\mathrm{rec},*}(\bar{F}_i^*)\bar{F}_i^* + w_{i,\mathrm{inh}}\bar{F}_{\mathrm{inh}}^*(\bar{F}_1^*, \bar{F}_2^*) + \sum_k \bar{n}^{\mathrm{ff}}\bar{w}_{ik}^{\mathrm{ff},*}(\bar{F}_i^*)\bar{I}_k\right).$$

We solve this system numerically to receive the equilibrium values $\bar{u}_1^*$ and $\bar{u}_2^*$ and, in consequence, by means of equations 10, 11 and 12, also $\bar{w}_1^{\mathrm{rec},*}$, $\bar{w}_2^{\mathrm{rec},*}$, $\bar{w}_{1\mathrm{A}}^{\mathrm{ff},*}$, $\bar{w}_{1\mathrm{B}}^{\mathrm{ff},*}$, $w_{2\mathrm{A}}^{\mathrm{ff},*}$, $\bar{w}_{2\mathrm{B}}^{\mathrm{ff},*}$ and $\bar{u}_{\mathrm{inh}}^*$.

Stability

The stability of an equilibrium is determined by the sign of the eigenvalue with the largest real part of the system's Jacobi matrix evaluated at the equilibrium. The nonzero terms of the Jacobi matrix are ($i \in \{1, 2\}$, $k \in \{\mathrm{A}, \mathrm{B}\}$):

$$\frac{\partial\dot{\bar{u}}_i}{\partial\bar{u}_i} = -\frac{1}{\tau} + R\bar{n}_i^{\mathrm{rec}}\bar{w}_i^{\mathrm{lat}}\frac{\partial\bar{F}_i}{\partial\bar{u}_i}, \qquad \frac{\partial\dot{\bar{u}}_i}{\partial\bar{u}_{\mathrm{inh}}} = Rw_{i,\mathrm{inh}}\frac{\partial\bar{F}_{\mathrm{inh}}}{\partial\bar{u}_{\mathrm{inh}}},$$

$$\frac{\partial\dot{\bar{u}}_i}{\partial\bar{w}_i^{\mathrm{rec}}} = R\bar{n}_i^{\mathrm{rec}}\bar{F}_i, \qquad \frac{\partial\dot{\bar{u}}_i}{\partial\bar{w}_{ik}^{\mathrm{ff}}} = R\bar{n}_i^{\mathrm{ff}}\bar{I}_k,$$

$$\frac{\partial\dot{\bar{u}}_{\mathrm{inh}}}{\partial\bar{u}_i} = R_{\mathrm{inh}}w_{\mathrm{inh},i}N_i\frac{\partial\bar{F}_i}{\partial\bar{u}_i}, \qquad \frac{\partial\dot{\bar{u}}_{\mathrm{inh}}}{\partial\bar{u}_{\mathrm{inh}}} = -\frac{1}{\tau_{\mathrm{inh}}},$$

$$\frac{\partial\dot{\bar{w}}_i^{\mathrm{rec}}}{\partial\bar{u}_i} = \mu^{\mathrm{rec}}\frac{\partial\bar{F}_i}{\partial\bar{u}_i}\left(2\bar{F}_i - \frac{(\bar{w}_i^{\mathrm{rec}})^2}{\kappa^{\mathrm{rec}}}\right), \qquad \frac{\partial\dot{\bar{w}}_i^{\mathrm{rec}}}{\partial\bar{w}_i^{\mathrm{rec}}} = \frac{2\mu^{\mathrm{rec}}}{\kappa^{\mathrm{rec}}}(F^{\mathrm{T}} - \bar{F}_i)\bar{w}_i^{\mathrm{rec}},$$

$$\frac{\partial\dot{\bar{w}}_{ik}^{\mathrm{ff}}}{\partial\bar{u}_i} = \mu^{\mathrm{ff}}\frac{\partial\bar{F}_i}{\partial\bar{u}_i}\left(\bar{I}_k - \frac{(\bar{w}_{ik}^{\mathrm{ff}})^2}{\kappa^{\mathrm{ff}}}\right), \qquad \frac{\partial\dot{\bar{w}}_{ik}^{\mathrm{ff}}}{\partial\bar{w}_{ik}^{\mathrm{ff}}} = \frac{2\mu^{\mathrm{ff}}}{\kappa^{\mathrm{ff}}}(F^{\mathrm{T}} - \bar{F}_i)\bar{w}_{ik}^{\mathrm{ff}}$$

with

$$\frac{\partial\bar{F}_i}{\partial\bar{u}_i} = \beta\bar{F}_i\left(1 - \frac{\bar{F}_i}{\alpha}\right) \qquad \text{and} \qquad \frac{\partial\bar{F}_{\mathrm{inh}}}{\partial\bar{u}_{\mathrm{inh}}} = \beta\bar{F}_{\mathrm{inh}}\left(1 - \frac{\bar{F}_{\mathrm{inh}}}{\alpha}\right).$$

The eigenvalues of the resulting matrix are determined numerically.

Feed-Forward Synaptic Weight Change

For constant pre- and postsynaptic activities (Figure 2 D), equation 11 in the main text can be solved analytically by separation of variables. The resulting time-course $w_i^{\mathrm{ff}}(t)$ depends on the given parameters and initial conditions:

$$\bar{w}_i^{\mathrm{ff}}(t) = \begin{cases} \bar{w}_i^{\mathrm{ff},*} \coth\left(\sqrt{\frac{\bar{F}_i \bar{I}_i (\bar{F}_i - F_{\mathrm{T}})}{\kappa^{\mathrm{ff}}}}(t-t_0)\mu^{\mathrm{ff}} + \operatorname{arcoth}\left(\frac{\bar{w}_i^{\mathrm{ff}}(t_0)}{w_i^{\mathrm{ff},*}}\right)\right) & \text{for } \bar{w}_i^{\mathrm{ff}}(t_0) > w_i^{\mathrm{ff},*} \wedge \bar{F}_i^{\mathrm{ff}} > F_{\mathrm{T}} \wedge \bar{I}_i > 0, \\ \bar{w}_i^{\mathrm{ff},*} \tanh\left(\sqrt{\frac{\bar{F}_i \bar{I}_i (\bar{F}_i - F_{\mathrm{T}})}{\kappa^{\mathrm{ff}}}}(t-t_0)\mu^{\mathrm{ff}} + \operatorname{artanh}\left(\frac{\bar{w}_i^{\mathrm{ff}}(t_0)}{w_i^{\mathrm{ff},*}}\right)\right) & \text{for } \bar{w}_i^{\mathrm{ff}}(t_0) < w_i^{\mathrm{ff},*} \wedge \bar{F}_i^{\mathrm{ff}} > F_{\mathrm{T}} \wedge \bar{I}_i > 0, \\ \bar{w}_i^{\mathrm{ff},\dagger} \tan\left(\sqrt{\frac{\bar{F}_i \bar{I}_i (\bar{F}_i - F_{\mathrm{T}})}{\kappa^{\mathrm{ff}}}}(t-t_0)\mu^{\mathrm{ff}} + \arctan\left(\frac{\bar{w}_i^{\mathrm{ff}}(t_0)}{w_i^{\mathrm{ff},*}}\right)\right) & \text{for } \bar{w}_i^{\mathrm{ff}}(t_0) < w_i^{\mathrm{ff},*} \wedge \bar{F}_i^{\mathrm{ff}} < F_{\mathrm{T}} \wedge \bar{I}_i > 0, \\ \left(\frac{1}{\bar{w}_i^{\mathrm{ff}}(t_0)} - \frac{F^{\mathrm{T}} - \bar{F}_i^{\mathrm{ff}}}{\kappa^{\mathrm{ff}}}(t-t_o)\mu^{\mathrm{ff}}\right)^{-1} & \text{for } \bar{F}_i^{\mathrm{ff}} = 0 \vee \bar{I}_i = 0 \end{cases}$$

with

$$\bar{w}_i^{\mathrm{ff},\dagger}(\bar{F}_i^*) = \sqrt{\frac{\kappa^{\mathrm{ff}} \bar{F}_i^* \bar{I}}{F^{\mathrm{T}} - \bar{F}_i^*}}.$$

Recruitment Basins

For determining the recruitment basins (Figure 2 E and Figure 3), we exploit the symmetry of the system and that, in general, only one of the two stimuli (A or B) is active. Accordingly, we approximate the second, inactive input to zero and neglect the respective feed-forward synapses. The population model is integrated with the given initial values of the feed-forward and recurrent weights and $\bar{u}_1 = \bar{u}_2 = \bar{u}_{\mathrm{inh}} = 0$ for 100 s. At $t = 100\,\mathrm{s}$, we evaluate which of the two populations is active. Table S1 provides the exact used initial values.

S 1. Initial values for Recruitment Basin Plots

Figure	$\bar{u}_1$	$\bar{u}_2$	$\bar{u}_{\mathrm{inh}}$	$\bar{w}_{1\mathrm{A}}^{\mathrm{ff}}$	$\bar{w}_{2\mathrm{A}}^{\mathrm{ff}}$	$\bar{w}_1^{\mathrm{rec}}$	$\bar{w}_2^{\mathrm{rec}}$
2E left; 2E right†; 3B top left, 3B top right†, 3B bottom right*†, 3D top left, 3D top right†	0	0	0	$0, \Delta w_{1\mathrm{A}}^{\mathrm{ff}}, \ldots, \hat{w}_{1\mathrm{A}}^{\mathrm{ff}}$	$0.35\,\hat{w}_{2\mathrm{A}}^{\mathrm{ff}}$	$0, \Delta w_1^{\mathrm{rec}}, \ldots, \hat{w}_1^{\mathrm{rec}}$	$0.25\,\hat{w}_2^{\mathrm{rec}}$
3B bottom left*	0	0	0	$0, \Delta w_{1\mathrm{A}}^{\mathrm{ff}}, \ldots, \hat{w}_{1\mathrm{A}}^{\mathrm{ff}}$	$0.40\,\hat{w}_{2\mathrm{A}}^{\mathrm{ff}}$	$0, \Delta w_1^{\mathrm{rec}}, \ldots, \hat{w}_1^{\mathrm{rec}}$	$0.25\,\hat{w}_2^{\mathrm{rec}}$
3D bottom left*	0	0	0	$0, \Delta w_{1\mathrm{A}}^{\mathrm{ff}}, \ldots, \hat{w}_{1\mathrm{A}}^{\mathrm{ff}}$	$1.00\,\hat{w}_{2\mathrm{A}}^{\mathrm{ff}}$	$0, \Delta w_1^{\mathrm{rec}}, \ldots, \hat{w}_1^{\mathrm{rec}}$	$1.00\,\hat{w}_2^{\mathrm{rec}}$
3D bottom right*†	0	0	0	$0, \Delta w_{1\mathrm{A}}^{\mathrm{ff}}, \ldots, \hat{w}_{1\mathrm{A}}^{\mathrm{ff}}$	$0.00\,\hat{w}_{2\mathrm{A}}^{\mathrm{ff}}$	$0, \Delta w_1^{\mathrm{rec}}, \ldots, \hat{w}_1^{\mathrm{rec}}$	$1.00\,\hat{w}_2^{\mathrm{rec}}$

$$\Delta w_{1\mathrm{A}}^{\mathrm{ff}} = 0.001 \hat{w}_{1\mathrm{A}}^{\mathrm{ff}},\ \Delta w_1^{\mathrm{rec}} = 0.001 \hat{w}_1^{\mathrm{rec}}$$

*) Using symmetry by commutating populations. †) Using symmetry by commutating inputs.

Comparison of Bifurcation Curve with Network Simulation

When comparing the equilibrium structure of the population model dependent on the input amplitude (bifurcation parameter) with the equilibria reached in network simulations (Figure 2 C), the network simulations are initialized close to the different expected stable configurations. For every input amplitude I, we perform two simulations with different initial conditions:

- $w_{ij}^{\mathrm{rec}} = 0.25\hat{w}^{\mathrm{rec}}$ for all realized recurrent synapses and $w_{ij}^{\mathrm{ff}} = \hat{w}^{\mathrm{ff}}$ for all realized feed-forward synapses.
- $w_{ij}^{\mathrm{rec}} = \hat{w}^{\mathrm{rec}}$ for synapses in between 121 neurons in a circle-shaped population, $w_{ij}^{\mathrm{rec}} = 0.25\hat{w}^{\mathrm{rec}}$ for all other realized recurrent synapses and $w_{ij}^{\mathrm{ff}} = \hat{w}^{\mathrm{ff}}$ for all realized feed-forward synapses.

In each case, the network is simulated for $50,000\,\mathrm{s}$. Every simulation is repeated 50 times with different random connectivities. To avoid simulation artifacts related to absolute silence of input channels, we assume a small background activity of 0.1α for inactive inputs. In the final state, we either consider all neurons with activity higher than 0.5α or, if there are none, 120 neurons centered around the activity center of the network as population 1. Population 2 is defined as the circular group of 120 neurons with the highest distance (respecting the periodic boundary conditions) to population 1. Within these two population, we evaluate the mean recurrent weight.

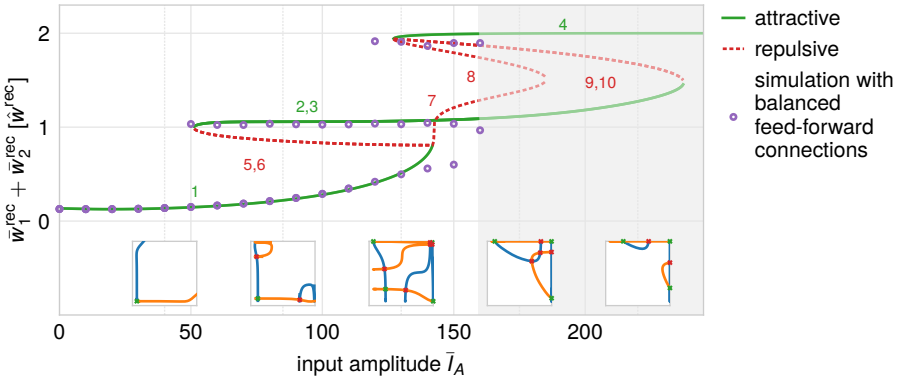
Large Input Amplitudes In the network simulation, the functional role of the inhibitory population is two-fold: On the one hand, inhibition mediates the competition between different populations. This role is also captured by the population model. On the other hand, it prevents an active cell assembly from growing without limit by inhibiting neighboring neurons. This aspect is not reflected in the population model as in the latter the size of the populations is approximated as being fixed. Due to this discrepancy, the population model predicts equilibria also for very large input amplitudes while in the network simulation these input amplitudes lead to full activation of the complete network. The respective area is shown in gray in Figure 2 C and Figure S6.

Symmetrized Network Simulation As described before, the population model differs from the complete network simulation by not capturing the two-fold role of the inhibitory population. Additionally, as a mean-field-like approach, it cannot reflect small symmetry breaks due to the random feed-forward connectivity. In combination, these two constraints are the source of a deviation of the population model from the results of the network simulation: In the latter, simultaneous activity of two distinct populations is not observed. This is due to the fact that the exact amplitude of inhibition, needed to balance the recurrent excitation and thereby to stop the growth of a given cell assembly, depends on the random feed-forward connectivity and is therefore different at every network location. As a result, the attractive equilibrium 4 is never reached (Figure 2 C).

To illustrate that this is indeed due to the heterogeneity introduced by the random feed-forward connectivity, we perform another set of network simulations with balanced feed-forward connections, i.e. we make sure that every neuron receives exactly two synapses from each of the two input populations. Apart from $\bar{n}^{\mathrm{ff}} = 2$, this also slightly influences the other cell assembly topology parameters where we now find $N \approx 121$ and $\bar{n}^{\mathrm{rec}} \approx 34.9$. For every tested input amplitude I, we perform simulations with three different initial conditions:

- $w_{ij}^{\mathrm{rec}} = 0.25\hat{w}^{\mathrm{rec}}$ for all realized recurrent synapses and $w_{ij}^{\mathrm{ff}} = \hat{w}^{\mathrm{ff}}$ for all realized feed-forward synapses.
- $w_{ij}^{\mathrm{rec}} = \hat{w}^{\mathrm{rec}}$ for synapses in between 121 neurons in a circle-shaped population, $w_{ij}^{\mathrm{rec}} = 0.25\hat{w}^{\mathrm{rec}}$ for all other realized recurrent synapses and $w_{ij}^{\mathrm{ff}} = \hat{w}^{\mathrm{ff}}$ for all realized feed-forward synapses.
- $w_{ij}^{\mathrm{rec}} = \hat{w}^{\mathrm{rec}}$ for synapses in between each of the 121 neurons in two circle-shaped populations, $w_{ij}^{\mathrm{rec}} = 0.25\hat{w}^{\mathrm{rec}}$ for all realized recurrent synapses and $w_{ij}^{\mathrm{ff}} = \hat{w}^{\mathrm{ff}}$ for all realized feed-forward.

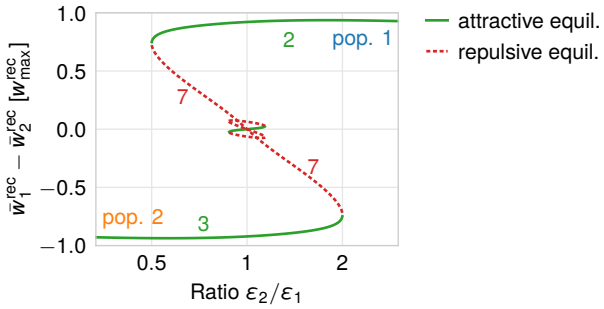
As expected, in this balanced configuration the attractive state 4 representing two simultaneously active populations is reached (Figure S6).



S 6. Input bifurcation diagram of the network with balanced feed-forward connections. In contrast to Figure 2 C in the main text, an attractive state with two active populations is reached in the network simulation with balanced feed-forward connections. The gray area indicates input amplitudes $\bar{I}_A$ for which the inhibition is no longer able to circumvent unlimited growth of cell assemblies.

Excitability Ratio Bifurcation Diagram

In Figure 4 B in the main text, we only show a schematic version of the bifurcation curve as a function of the excitabilities of the two populations. Figure S7 shows the exact curve as obtained from the population model.



S 7. Excitability bifurcation diagram as obtained from the population model. Note that a higher value of ϵ_i means a lower excitability of population i and vice-versa. For a given ratio $\frac{\epsilon_2}{\epsilon_1}$, the used values are $\epsilon_1 = \epsilon_0 \cdot \sqrt{\frac{\epsilon_2}{\epsilon_1}}^{-1}$ and $\epsilon_2 = \epsilon_0 \cdot \sqrt{\frac{\epsilon_2}{\epsilon_1}}$.