

## **Supplemental Information**

### **Neuronal Circuits Underlying**

### **Persistent Representations**

### **Despite Time Varying Activity**

Shaul Druckmann and Dmitri B. Chklovskii

#### **Supplemental Experimental Procedures**

##### *Construction of sparse FEVER networks*

The FEVER rule is expressed by the equation:

$$\vec{d}_i = \sum_j \vec{d}_j L_{ji}, \text{ for all } 1 \leq i \leq n \quad \text{Eq. M1}$$

where the vectors  $d$  are feature vectors and  $L$  is the matrix of lateral connectivity. As stated in the main text, this equation can be interpreted as recombining each feature vector by a linear combination of the other feature vectors.

In this section we describe constructing more biologically realistic FEVER networks satisfying Dale's law by decomposing the connectivity matrix into a sum of two components: a sparse matrix of excitatory-excitatory interactions ( $E$ ) and a low rank, non-sparse, matrix of inhibitory interactions ( $N$ ) due to disynaptic interactions between excitatory cells via inhibitory neurons. The low rank of inhibitory interactions reflects the fact that these interactions need to be mediated by a much smaller (~20%) population of inhibitory interneurons.

Thus, the cost function to be minimized contains three terms: the reconstruction error which measures the square deviation between the feature vectors and their recombination, a term that penalizes the total strength of excitatory interactions,  $E$ , and a term that penalizes according to the convex relaxation of the rank of  $N$ , the inhibitory interactions:

$$C = \left( \vec{d}_i - \sum_j \vec{d}_j (E_{ji} - N_{ji}) \right)^2 + \lambda_1 \sum_i \|\vec{E}_i\|_1 + \lambda_2 \|N\|_* \quad \text{Eq. M2}$$

where  $\|x\|_*$  indicates the nuclear norm of  $x$ . This cost function was minimized subject to the constraint of the entries of  $E$  and  $N$  being non-negative. Minimizing this cost function with the appropriate constraints was performed by an iterative thresholding approach [1] implemented in Matlab. At each step of the iteration the full correction to the two matrices is calculated, which would maximally reduce the cost function. However, instead of making this correction, a modified correction is applied subject to two thresholding operations. First, the rank of the inhibitory (negative) synaptic weights matrix undergoes the threshold operation by retaining only the contributions from singular values greater than threshold. If the matrix is given by  $W=U*S*V'$ , this is performed by setting the diagonal of the singular value matrix,  $S$ , to zero for values less than the threshold value and then reconstructing  $W$ . Second, the excitatory connections are subject to a shrinkage operation, namely reducing all values by the threshold value and setting to zero those whose absolute value was less than threshold. Full details of the process by which these operations are related to the cost function can be found in the reference [1].

The values of the two Lagrange multipliers ( $\lambda_1, \lambda_2$ ) were set so as to maintain high accuracy while matching the properties of  $E$  and  $N$  to that of cortical networks, i.e., approximately 5% connectivity and approximately 20% inhibitory interneurons (rank). Matrices obtained for these values are those whose dynamics are portrayed in Figure 3. Reasonable changes in the connectivity and rank values (perturbing in a range equal to 20% of their value) did not result in qualitative changes to the results presented in this study.

#### *Persistent network stimulus representations and time varying activity*

Let us consider the eigenvectors of the connectivity matrix,  $L$ . Denoting the  $i$ th right eigenvector of  $L$  by  $v_i$ , and the corresponding eigenvalue by  $\lambda_i$ , i.e.,  $L\vec{v}_i = \lambda_i\vec{v}_i$ , and multiplying both sides of Eq. 5 (rewritten in matrix format to include all feature vectors simultaneously) by  $\vec{v}_i$  we obtain:

$$D\vec{v}_i = DL\vec{v}_i, \text{ or}$$

$$D\vec{v}_i = \lambda_i D\vec{v}_i \quad \text{Eq. M3}$$

To satisfy this equation at least one of the following conditions must be met:

$$\lambda_i = 1$$

$$D\vec{v}_i = 0. \quad \text{Eq. M4}$$

In other words, if the eigenvector  $v_i$  is not in the null-space of  $D$  it must have unitary eigenvalue and therefore be persistent. If the eigenvector  $v_i$  is in the null-space it may correspond to an arbitrary eigenvalue, thus allowing for time-varying activity.

One can also relate the independence of the feature vectors to the time dependence of activity more directly. First we show that if the vectors  $d_i$  are linearly independent the network stimulus representation is persistent for all stimuli,  $s$ , if and only if the activity,  $a$ , is constant in time. Re-expressing the derivative of the network stimulus representation in terms of activity we obtain:

$$\frac{d\vec{s}}{dt} = \frac{d}{dt} \left( \sum_i \vec{d}_i a_i \right) = \sum_i \vec{d}_i \frac{d}{dt} a_i \quad \text{Eq. M5}$$

To prove the sufficiency of the constant activity, note that, if the derivative of all components of  $a$  is zero, then the derivative of  $s$  is zero and the network stimulus representation remains constant.

To prove the necessity of the constant activity condition, assume for the purpose of contradiction, that persistent network stimulus representations can be maintained with time-varying activity. Thus, there are at least two different patterns of neuronal activity,  $a^1$  and  $a^2$ , which encode the same stimulus,  $s$ :

$$\vec{s} = \sum_i \vec{d}_i a_i^1 \quad ; \quad \vec{s} = \sum_i \vec{d}_i a_i^2 \quad ; \quad a_i^1 \neq a_i^2 \quad \text{Eq. M6}$$

where the superscript denotes the identity of the pattern (1 or 2) and the subscript denotes the summing index.

Now consider the difference of the two representations afforded by the two activity patterns:

$$\begin{aligned} \sum_i \vec{d}_i a_i^1 - \sum_i \vec{d}_i a_i^2 &= \vec{s} - \vec{s} = 0 \\ \sum_i \vec{d}_i a_i^1 - \sum_i \vec{d}_i a_i^2 &= \sum_i \vec{d}_i (a_i^1 - a_i^2) \equiv \sum_i \vec{d}_i a_i^3 \end{aligned} \quad \text{Eq. M7}$$

where we defined  $a^3$  as the difference between the two patterns. Combining the two equations above we find:

$$\sum_i \vec{d}_i a_i^3 = 0 \quad \text{Eq. M8}$$

contradicting the assumption that the vectors  $\vec{d}_i$  are independent since we found a linear sum of them with non-zero coefficients that adds up to zero.

In a linearly dependent representation framework, persistent network stimulus representations can be achieved with time variant activity. To see this, recall that linear dependence implies that, for some  $k$ ,  $\vec{d}_k$  can be expressed as:

$$\vec{d}_k = \sum_{i \neq k} c_i \vec{d}_i \Rightarrow \sum_i \vec{d}_i \cdot (c_1 \ c_2 \ \dots \ c_{k-1} \ -1 \ c_{k+1} \ \dots \ c_n) = \vec{d}_k - \vec{d}_k = 0 \quad \text{Eq. M9}$$

If we view the set of coefficients in the parenthesis as a direction in activity space, then the change in activity along this direction produces no change in the network stimulus representation  $\vec{s} = \sum \vec{d}_i a_i$ . Therefore, in at least one direction in activity space, time variant activity results in time invariant network stimulus representations.

#### *Stable FEVER networks cannot contain only excitatory interactions*

We prove that a network that has only excitatory interactions constructed according to the FEVER rule (and that does not contain only autapses) will have at least one eigenvalue greater than one indicating that its dynamics are unstable. The Perron-Frobenius theorem for an irreducible, non-negative matrix  $A$  with spectral radius  $r$ , states (among other properties) that the number  $r$  is a positive real number and there exists an eigenvalue of  $A$  equal to  $r$ . The spectral radius for such a matrix obeys the following inequality:

$$\min_j \sum_i a_{ij} \leq r \leq \max_j \sum_i a_{ij} \quad \text{Eq. M10}$$

Consider the minimum value of the sum of the different columns. Equation M1 states that the jth column of L corresponds to the coefficients of a sum over all feature vectors (besides the jth) that equals the feature vector of the jth neuron. The triangle inequality:

$$\|x + y\| \leq \|x\| + \|y\| \quad \text{Eq. M11}$$

states that the norm of the sum of two vectors is at most as large as the sum of the norms of the two vectors. For the sake of simplicity let us assume that for all neurons the FEVER principle can be accomplished by a sum over only two other neurons. Without loss of generality let us look at the connection from the first neuron to the two neurons that recombine its feature vectors, neurons two and three:

$$\vec{d}_1 = \vec{d}_2 L_{21} + \vec{d}_3 L_{31} \quad \text{Eq. M12}$$

Since all the feature vectors are normalized to the same value (one), we can use the triangle inequality and obtain:

$$1 \leq \|\vec{d}_2 L_{21}\| + \|\vec{d}_3 L_{31}\| = \|L_{21}\| + \|L_{31}\| \quad \text{Eq. M13}$$

Thus the square of the sum of (non-negative) elements of the 1<sup>st</sup> column of L is greater than one. This holds true for all neurons and is not dependent on the recombination being obtained from only two neurons, as the triangle inequality can also be applied recursively. Equality will be obtained only if one of the two neurons' feature vector is exactly equal to the first neuron's feature vector in every aspect (or technically equal to minus the feature vector). We do not consider this artificial case (or consider simply the network were we discard that neuron). Thus, the minimum of the sum of the columns of L is greater than one. Accordingly the spectral radius of the matrix is greater than one and there will be at least one eigenvalue greater than one, indicating that the network dynamics to be unstable.

We note that the above calculations bear two important exceptions. First the Perron-Frobenius theorem applies only to irreducible matrices (a matrix is reducible if and only if for some permutation matrix P the matrix P<sup>T</sup>AP is block upper triangular). Secondly, we assumed that the feature vectors are not all

identical up to scaling. This exempts the important case of scalar feature vectors (that are by definition identical to each other up to a scaling factor).

### *Hebbian learning of FEVER networks with simple feature vectors*

As described in the main text we consider Hebbian (correlation based) learning of FEVER networks in the case of the set of feature vectors, expressed in the matrix,  $D$ , forming a tight frame [2].

Since  $D$  is a tight frame then:

$$DD^T = \lambda I \quad \text{Eq. M14}$$

where  $I$  is the identity matrix and  $\lambda > 1$ . Consider a lateral connectivity network proportional to the correlation of the feature vectors with no autapses:

$$L = \frac{1}{(\lambda-1)}(D^T D - I) \quad \text{Eq. M15}$$

Let us show that this network is indeed a FEVER network:

$$DL = \frac{D}{(\lambda-1)}(D^T D - I) \quad \text{Eq. M16}$$

Plugging in the condition of the tight frame:

$$DL = \frac{(\lambda-1)D}{(\lambda-1)}, \\ DL = D \quad \text{Eq. M17}$$

Which is the FEVER rule.

The Hebbian learning rule is defined as:

$$L_{ij}^{new} = L_{ij}^{old} + \eta a_i a_j, \quad \text{Eq. M18}$$

where  $\eta$  is the learning rate and  $a_i$  the activity of the  $i$ th neuron.

Consider a feedforward network from an input layer of  $m$  neurons, to a layer of  $n$  neurons. The feature vectors,  $D$ , of the  $n$  neurons are defined by the feedforward weights,  $D^T$ . We consider brief transient input followed by the delay activity period. In other words we assume the initial activity (denoted by  $a_0$ ) to be equal to the projection of the lower dimensional stimulus through the feedforward weights:

$$a_0 = D^T s \quad \text{Eq. M19}$$

We consider the low dimensional input to the network to be whitened, since it is well known that circuits prior to cortex perform whitening (e.g., retina and LGN for visual cortex [3]). This is expressed by:

$$\langle ss^T \rangle = I, \quad \text{Eq. M20}$$

where the angular brackets denote averaging across different presentations of stimuli.

We begin with the lateral connectivity matrix at zero and therefore the activity is described by the solution of  $\frac{da}{dt} = -a$ :

$$a(t) = \exp(-t) D^T s \quad \text{Eq. M21}$$

A Hebbian learning rule will result in the non-diagonal elements of  $L$  evolving according to:

$$\Delta L = \eta \exp(-2t) D^T ss^T D \quad \text{Eq. M22}$$

We note here that learning taking place at different times during the dynamics just results in a different effective learning rate, which we incorporate into the actual learning rate in what follows to avoid clutter.

Assuming that  $\eta$  is small we can approximate any individual contribution to the synaptic weights by an average over such steps:

$$\begin{aligned} \Delta L &= \eta D^T \langle ss^T \rangle D \\ \Delta L &= \eta D^T D \end{aligned} \quad \text{Eq. M23}$$

Having plugged in the whitening assumption, Eq. M20. Thus, we find that at least while the contribution of the lateral connections to the dynamics can be neglected, the synaptic matrix changes in the direction of  $D^T D$ , which we recall is a FEVER network (Eq. M17).

At some point we can no longer neglect the contribution of the lateral connections to the dynamics. Assuming that this point is reached after a certain number of learning iterations we have:

$$L = \beta D^T D, \quad \text{Eq. M24}$$

where  $\beta$  incorporates both the learning rate and the number of iterations that have taken place. The dynamics now are given by:

$$\frac{da}{dt} = -a + La,$$

$$\frac{da}{dt} = -(\beta(D^T D - I) - I)a$$

and is solved as above by:

$$a(t) = \exp(-((\lambda - 1)\beta - 1)t) D^T s \quad \text{Eq. M25}$$

Accordingly the non-diagonal elements of L will evolve in the following manner:

$$\Delta L = \eta \exp(-2((\lambda - 1)\beta - 1)t) D^T \langle ss^T \rangle D \quad \text{Eq. M26}$$

Therefore the non-diagonal elements stay proportional to  $D^T D$ , which obeys the FEVER rule.

In summary, we show that a Hebbian learning rule results in a matrix update that is proportional to a FEVER matrix in early, intermediate and late times. Hebbian learning rules are well known to be unstable, leading to unbounded growth of the synaptic weights [4]. We therefore adopt the common approach of adding a divisive normalization constraint to the learning [4]. To allow for the extension of different time scales of population activity (i.e.,  $\alpha \neq 1$  in main text equation 6) one also needs to add to the learning rule a simple, global scalar learning signal that indicates whether the activity has been sustained for a long enough period of time. Such a global signal can be mediated by feedback activity from a downstream network or neurotransmitter release. We note that this learning scenario is clearly abstract and simplistic. It is meant to show that the FEVER architecture can be learned by a biologically accepted learning rule, not to offer a detailed picture of true synaptic learning which is known to be considerably more complex.

A particularly simple and interesting example of a tight frame in the two-dimensional case is a set of feature vectors with  $n$  neurons equal to one original feature vector and  $(n-1)$  incremental rotations by the angle:  $\alpha = (2\pi)/(n)$ , where  $n$  is an odd number. Thus, the angle between any two neurons is  $(2\pi/n)$ . Since all the feature vectors are unit vectors, the correlation between the feature vectors is simply cosine of the angle between them. Thus the feature vector correlation matrix  $D^*D$  is a square circulant matrix of dimension  $n$  whose first row is equal to:

$$[1; \cos(\alpha); \cos(2\alpha) \dots \cos((n-1)\alpha)] \quad \text{Eq. M27}$$

Let us show that this correlation yields a FEVER network, i.e., that:

$$D = DL \quad \text{Eq. M28}$$

where  $L = D^T D$  after setting the diagonal elements (corresponding to autapses) to zero.

Without loss of generality let us assume the first neuron's feature vector points up the vertical direction. Now let us consider the rest of the neurons as pairs, the first a rotation of  $\pm\alpha$ , the second pair a rotation of  $\pm 2\alpha$  and so on. All in all there are  $(n - 1)/2$  such pairs. Recalling that the FEVER rule can be interpreted as a sum over all postsynaptic weights times their feature vectors should be equal to the neuron's own feature vector, let us now add up these contributions by the pairs defined above. Adding up the first pair, we find that their projection on the horizontal direction is of identical magnitude but opposite signs and therefore cancels out. The projection of each of this pair of vectors on the vertical direction is equal to  $\cos(\alpha)$ . The entries in  $L$  as defined above are also equal to the angle between the vectors and the original feature vector, i.e.,  $\cos(\alpha)$ . Thus, this pair contributed  $2(\cos(\alpha))^2$  in the vertical direction and nothing in the horizontal direction. The same logic applies to the next pair which will contribute  $2(\cos(2\alpha))^2$  in the vertical direction and nothing in the horizontal direction. Summing up across all pairs we have:

$$\sum_{i=1}^n L_{ij} d_i = d_j \left( \sum_{w=1}^{(n-1)/2} 2(\cos(w\alpha))^2 \right)$$

$$\sum_{i=1}^n L_{ij} d_i = d_j \left( \frac{n}{2} - 1 \right) \quad \text{Eq. M29}$$

Thus, the FEVER rule is met for this feature vector with a factor  $n/2-1$ .

By the symmetry of the set of the feature vectors the calculation for any of the other feature vectors will be the exact same calculation only on a rotated original feature vector. Thus, the FEVER rule is met for such sets of feature vectors. In other words, if the final result of learning is a matrix that is proportional to the correlation of feature vectors, the result will be a FEVER matrix.

We numerically validated the above calculations by simulating a network as described above. Namely, we draw random stimuli and simulate the rate dynamics

of the network for 2D feature vectors ( $m=2$ ) for different numbers of neurons ( $n=5,7,25$ ) and for full persistence. There are two parameters involved,  $\eta$ , the learning rate described above and a parameter that determines how many times during the dynamics do we apply the learning rule. As expected from the calculations above, we find that the results are insensitive to both parameters, as long as the learning rate is small enough, as is commonly found in learning studies. For all of the networks we find that the learning indeed converges to the FEVER solution as expected. In addition, we tested the stability of the learning rule to both non-zero initial conditions, and to random perturbations during the learning. We find that the network is stable even to perturbations of magnitude 10-20% of the strength of the weights.

We note that the synaptic connectivity matrix as defined above will be dense. However, the symmetry of the rotational feature vectors also allows for sparse matrices FEVER to be learned if one simply assumes that weak correlations do not yield synaptic strengthening, e.g., a threshold on the Hebbian learning rule. Intuitively, such a threshold reduces the number of pairs of rotated feature vectors that participate in the FEVER network. Since the FEVER network can be built by any number of pairs, as long as the threshold is smaller than the largest correlation between the feature vectors a FEVER network will be learned (otherwise no learning will occur). Different values of the threshold affect the eigenvalues of the resulting FEVER network (Fig. S5). The details of all possible forms of FEVER networks that can thus be learned are beyond the scope of the analysis presented here.

#### *FEVER networks as a high-dimensional integrator of ongoing stimuli*

In the main text we described the response of FEVER networks to transient stimuli. In the context of ongoing stimulus presentation FEVER networks acts as a high-dimensional integrator. Assuming that the network receives time-varying input  $f$ , we can calculate the evolution of the stimulus representation:

$$\tau \frac{da_i}{dt} = -a_i + \sum_j L_{ij}a_j + f_i \Rightarrow \tau \sum_i \vec{d}_i \frac{da_i}{dt} = \sum_i \vec{d}_i \left( -a_i + \sum_j L_{ij}a_j \right) + \sum_i \vec{d}_i f_i \Rightarrow \tau \frac{d\vec{s}}{dt} = \sum_i \vec{d}_i f_i$$

$$\vec{s} = \frac{1}{\tau} \int \sum_i \vec{d}_i f_i dt \quad \text{Eq. M30}$$

Intuitively, this result can be understood by thinking of a constant stream of stimuli as a sum of many transient stimuli spaced in time. If the network makes each transient stimulus persistent, as does the FEVER network, then each of the individual transient stimuli will remain and they will all be added up, or integrated, in the network stimulus representation.

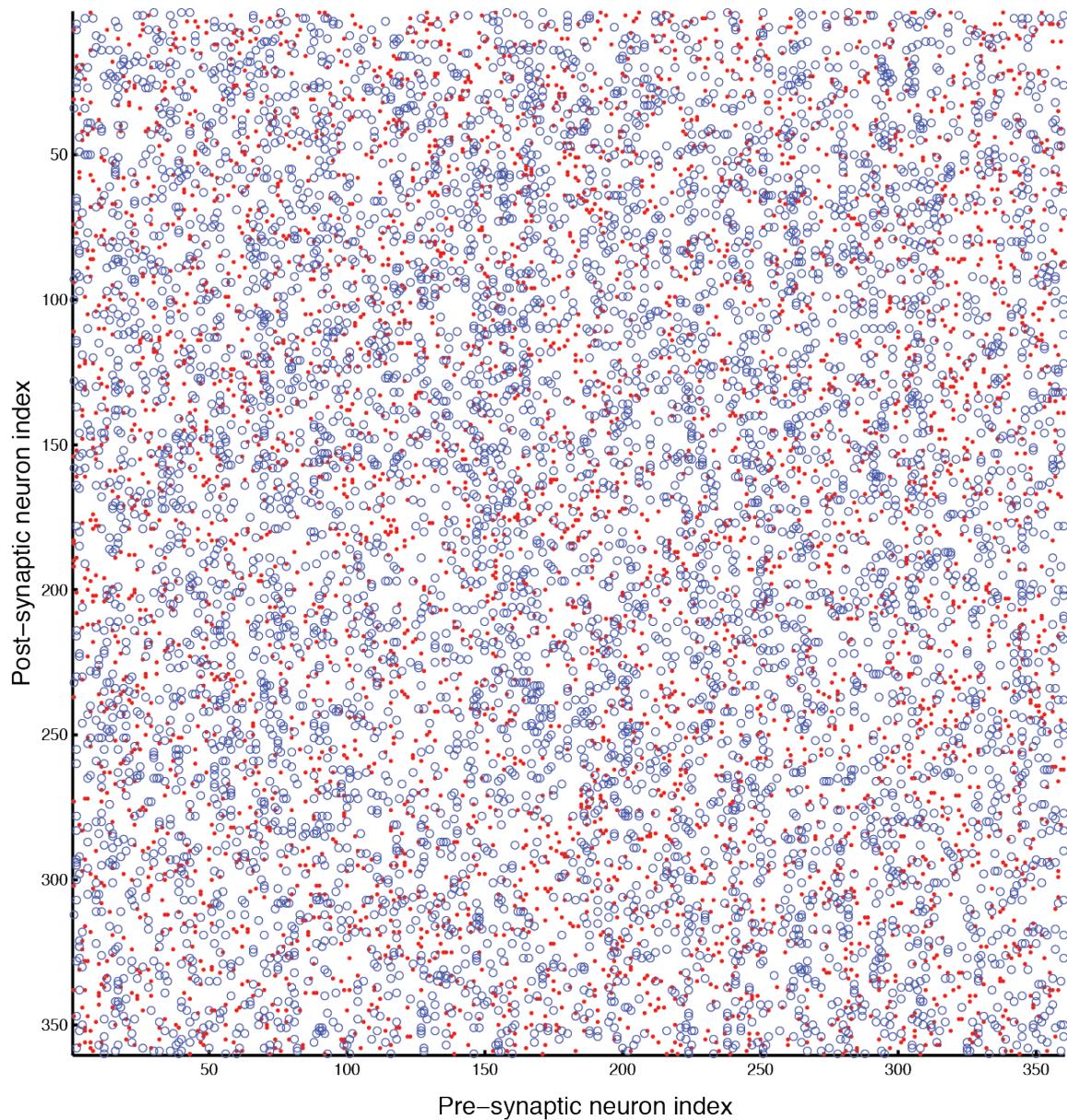
#### *Experimental testing of coding versus non-coding directions in activity space*

The most experimentally salient prediction regarding activity in FEVER networks is that there exist noncoding directions in activity space, corresponding to the nullspace of D, along with the coding directions and that activity during delay periods should be restricted to the noncoding directions. In order to experimentally find this difference without knowing D, one can attempt to extract which directions are coding and noncoding from activity. How can a coding direction be discovered? A straightforward approach would be to systematically vary one dimension of the stimulus and attempt to discover which direction in activity space this corresponds to. To evaluate such an attempt, we simulate the dynamics of a FEVER network, following multiple presentations of stimuli differing in value along one dimension. We emulate the problem of experimental subsampling of neurons by examining not the full activity simulated for the network, but the activity of subsets of the full network corresponding to one neuron, two neurons etc. For each of these subsets of recordings, we extract the coding direction by correlating the activity of the neurons with the different stimulus values. The weights of the coefficients define the coding direction in activity space of dimensionality corresponding to the size of the subsampled subset of neurons.

We empirically estimate the dimensions in which dynamics occur during delay periods by recording a second set of activity during delay periods and performing principal component analysis (PCA). We consider only the top 15% of the PCs (that capture approximately 90% of the variance) to avoid including the coding dimensions. Once we have estimated coding and noncoding directions in activity space we can determine whether activity is limited to the nullspace during delay activity by measuring the projection of the coding dimension on the delay activity subspace found by PCA (Figure S6). Simulation was performed using the same network as in Figure 3 and data from 1000 trials was generated and analyzed.

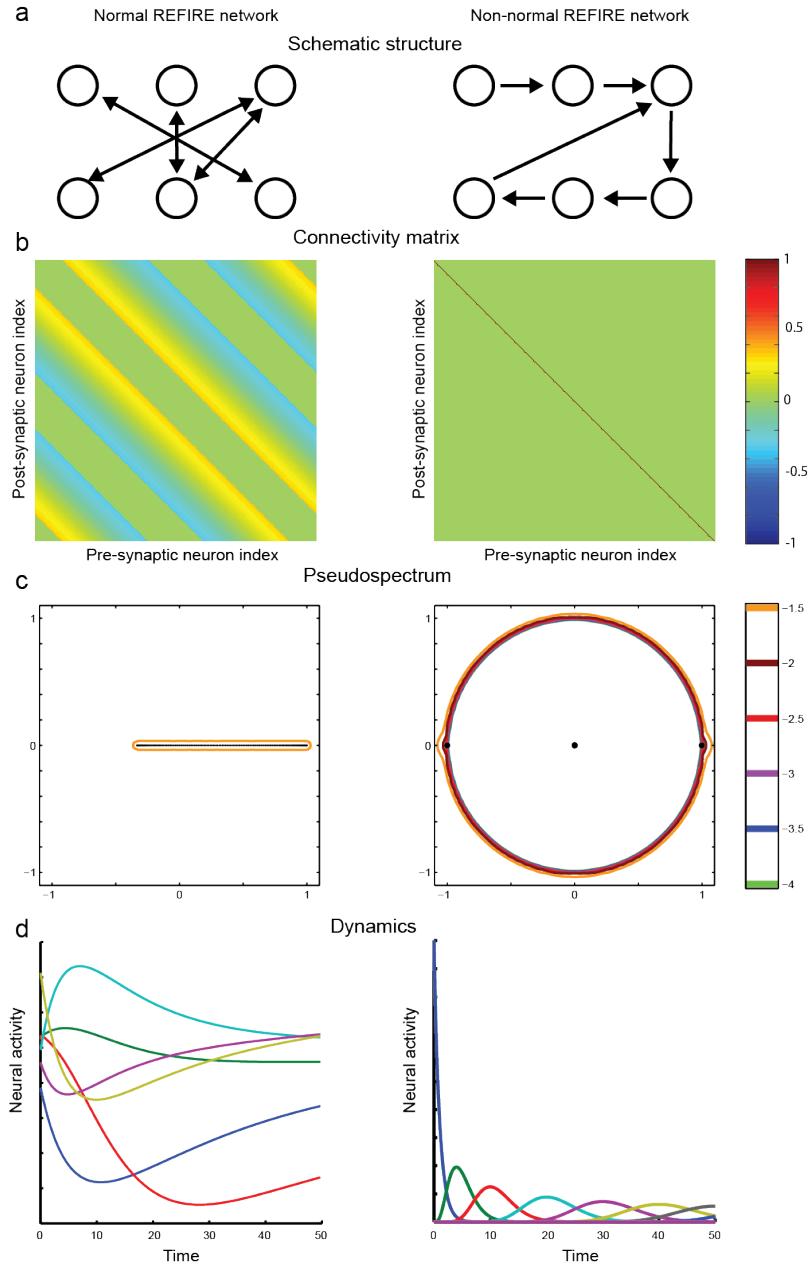
#### *Simulating integrate and fire networks of neural activity*

Networks of non-leaky (infinite time constant) integrate and fire neurons were simulated with the same connectivity rule. The resting and reset value of the integrate-and-fire neurons was set to 0 and the threshold to 1. This range would correspond roughly to -70mV through -50mV in cortical neurons. Synaptic transmission was modeled by simple current, not conductance, synapses, with an instantaneous rise and a decay of 1 ms (ten times the numerical simulation time step). The network was numerically simulated through the same integration algorithm as the rate model. The contribution of each neuron to the network stimulus representation is taken as the subthreshold activity times the neuron's feature vector.



**Figure S1. Network connectivity, Related to Figure 1**

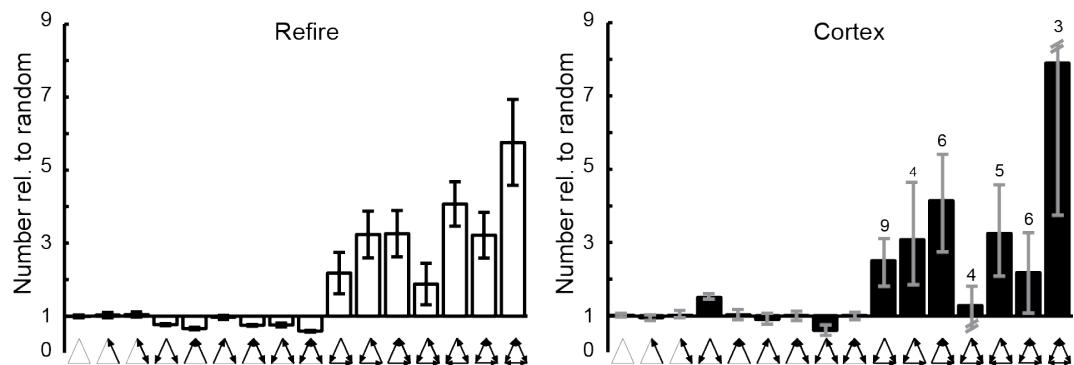
The synaptic connectivity between excitatory neurons is plotted with excitatory connections as red dots and inhibitory disynaptic connections as blue circles.



**Figure S2. FEVER networks can have both normal and non-normal dynamics, Related to Figure 2**

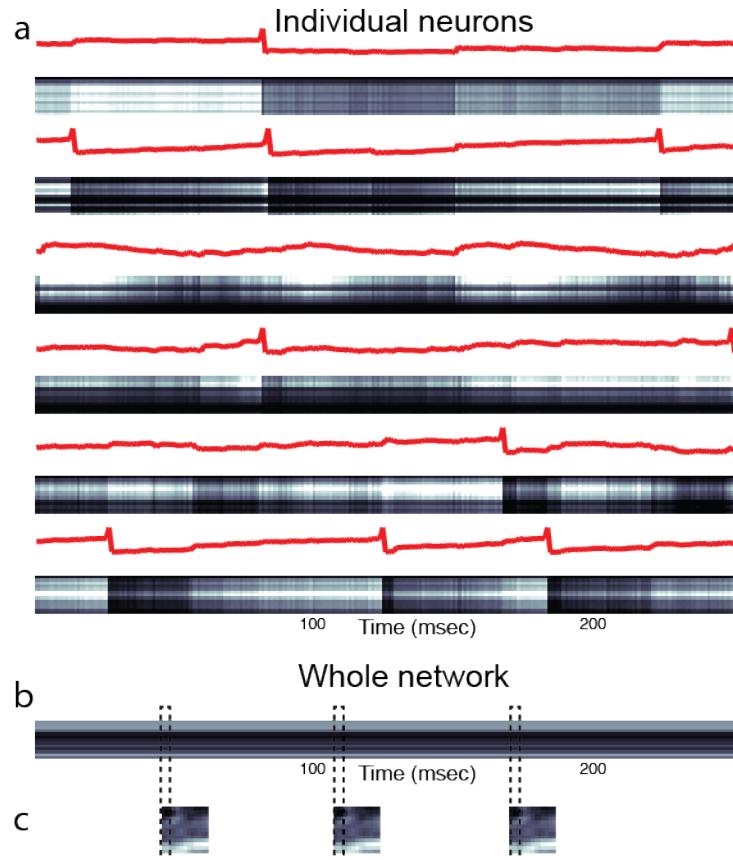
- a.** Schematic illustration of the connectivity matrix plotted for two FEVER networks, a normal (symmetric) FEVER network (left) and a non-normal FEVER matrix (right).
- b.** Color plot of connectivity matrix for both networks, bar on the right shows color code. **c.**  $\epsilon$ -pseudoeigenvalue spectrum of the networks. Eigenvalues of the networks are shown as black circles. Contours depict the boundaries of  $\epsilon$ -pseudoeigenvalues that satisfy the inequality  $\|(z - L)^{-1}\| > \epsilon^{-1}$  for different values of  $\epsilon$ . Bar on the far

right shows the color code for the different values of  $\epsilon$  in a log 10 basis. For the normal FEVER network, left, the pseudoeigenspectrum is simply the union of circles around the eigenvalues. For the non-normal FEVER network (right) the contours, circles in the plain, are very far from the eigenvalues which are on the real axis at -1,0 and 1. **d.** Activity traces from a subset of neurons for each network. Note that in both cases the variability far exceeds the time scale of a single membrane time constant.



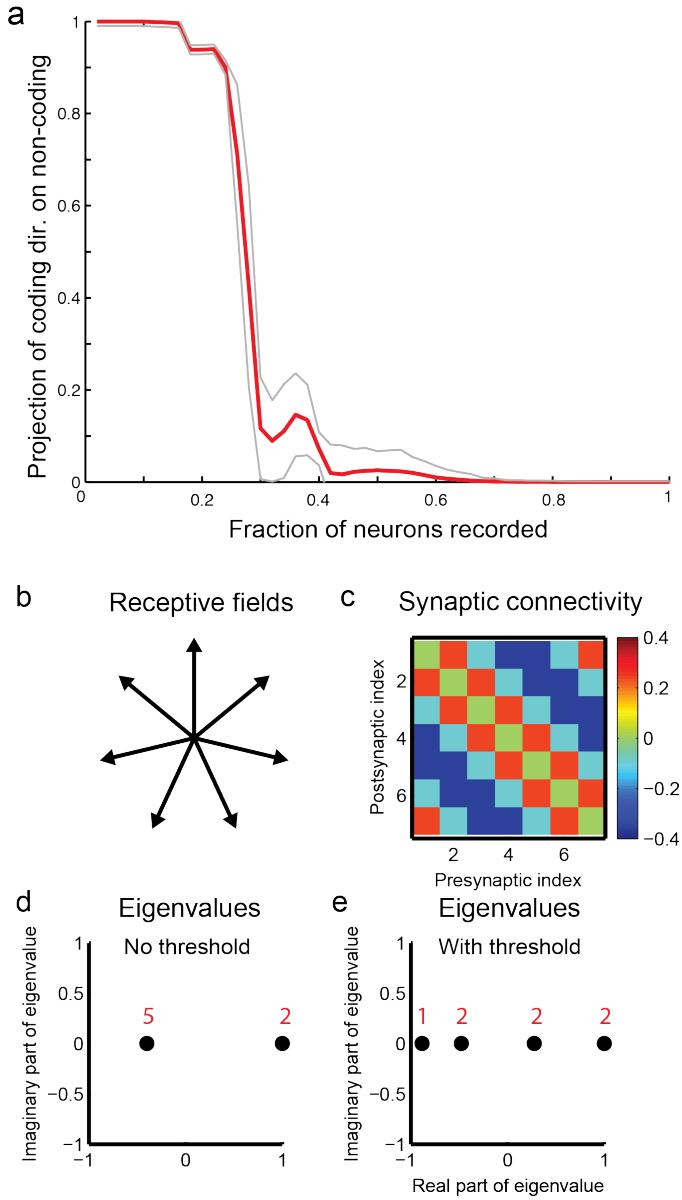
**Figure S3. Detailed triplet motif patterns show highly connected motifs to be overrepresented, Related to Figure 4**

Number of triplet motifs relative to that expected by random from a degree matched network for all triplet motifs for Cortex (right) and FEVER (left). Error bars indicate standard deviation. Numbers above fully connected triplet motifs correspond to number of actual motifs found.



**Figure S4. Spiking FEVER network, Related to Figure 3**

**a.** Red traces show membrane voltage for each of five neurons taken out of the integrate-and-fire FEVER network. Gray scale values underneath each trace show the contribution of that neuron to the network representation, i.e., the product of the neuron's sub-threshold potential and its feature vector, as a function of time. Only the first column out of the patch represented by each neuron is displayed to allow plotting as a function of time. Note that representation by each neuron constantly changes. **b.** Representation by full network. Note constancy of representation. **c.** Full patch represented by the network at three different times.



**Figure S5. FEVER network dynamics, Related to Figure 5**

**a.** Simulation of experimentally testing separation of activity into coding and non-coding dimensions. We empirically estimate coding and non-coding dimensions from simulated recordings from a FEVER network. Varying the fraction of recorded neurons from 0-1 we measure the projection of the estimated coding direction onto the non-coding space. A value of 0 indicates full orthogonality as expected from the FEVER model. Red line shows mean, gray lines indicate  $\pm$  one standard deviation. **b-e.** Learning simple FEVER networks **b.** Feature vector structure for simple two-dimensional stimulus representation. Arrows represent the direction in stimulus

space (i.e., the plane) of feature vector. **c.** Color map plot of learned synaptic connections. **d.** Eigenvalues of learned synaptic connections without threshold on learning rule. Red numbers above circles denote degeneracy of eigenvalue. Note that eigenvalue at real value of one has redundancy of two, the stimulus dimension. **e.** Eigenvalues of learned synaptic connections with threshold.

## **Supplemental References**

1. Cai, J.-F., Candes, E.J., and Shen, Z. (2010). A Singular Value Thresholding Algorithm for Matrix Completion. *SIAM Journal on Optimization* 20, 1956-1982.
2. Christensen, O. (2003). An introduction to frames and Riesz bases, (Boston: Birkhäuser).
3. Atick, J.J., and Redlich, A.N. (1993). Convergent Algorithm for Sensory Receptive-Field Development. *Neural Comput* 5, 45-60.
4. Miller, K.D., and Mackay, D.J.C. (1994). The Role of Constraints in Hebbian Learning. *Neural Comput* 6, 100-126.