

Journal Name

Crossmark

PAPER

RECEIVED
dd Month yyyyREVISED
dd Month yyyy

Daydreaming algorithm for Biased Patterns

Mikiya Doi^{1,*}, Masayuki Ohzeki^{1,2,3,4} and Federico Ricci-Tersenghi^{5,6,7,*}¹Graduate School of Information Sciences, Tohoku University, Sendai 980-8579, Japan²Department of Physics, Institute of Science Tokyo, Tokyo 152-8550, Japan³Research and Education Institute for Semiconductors and Informatics, Kumamoto University, Kumamoto 860-8555, Japan⁴Sigma-i Co., Ltd., Tokyo 108-0075, Japan⁵Physics Department, Sapienza University of Rome, P.le A. Moro 5, 00185 Rome, Italy⁶CNR–Nanotec, Rome unit, P.le A. Moro 5, 00185 Rome, Italy⁷INFN–Sezione di Roma 1, P.le A. Moro 5, 00185 Rome, Italy

*Authors to whom any correspondence should be addressed.

E-mail: mikiya.doi.t2@dc.tohoku.ac.jp, federico.ricci@uniroma1.it**Keywords:** Hopfield model, Biased patterns, Daydreaming algorithm**Abstract**

The *Daydreaming* algorithm has been recently proposed in Ref. [1] as a learning rule that simultaneously reinforces stored patterns and suppresses spurious attractors to improve the storage capacity of the Hopfield model. Its effectiveness has been reported for both uncorrelated and correlated data. However, the existing formulation has mainly assumed unbiased patterns, and the formulation for biased patterns has not yet been sufficiently established. Biased patterns are known to be much more problematic for models of associative memories.

In this study, we reformulate Daydreaming for biased patterns motivated by the underlying rationale of the pseudo-inverse rule. Specifically, we introduce the retrieval dynamics and an energy function based on the centered representation, and we derive a corresponding update rule for centered Daydreaming. We compare the centered pseudo-inverse rule with centered Daydreaming for biased patterns and examine the retrieval maps and eigenvalue distributions of the coupling matrices.

As a result, the centered Daydreaming yields larger basins of attraction than the centered pseudo-inverse rule, and such a beneficial property seems to be due to the broadness of the eigenvalue spectrum of the coupling matrix. To better understand this connection, we construct modified coupling matrices whose spectra interpolate between a pseudo-inverse-like spectrum and the Daydreaming one. The results clearly indicate that the broader spectrum generated by Daydreaming contributes to the enlarged basin of attraction.

1 Introduction

The Hopfield model is the best-known example of an attractor neural network [2, 3]. The model embeds desired memory patterns as stable fixed points of the retrieval dynamics by appropriately designing the coupling matrix $\mathbf{J}$. The original Hopfield model has neurons that take binary values $s_i \in \{+1, -1\}$ and evolve according to the following retrieval dynamics rule

$$s_i^{(t+1)} = \text{sign} \left(\sum_{j \neq i} J_{ij} s_j^{(t)} \right). \quad (1)$$

If the initial state lies within the basin of attraction of a stored pattern, the retrieval dynamics converge to that pattern. Therefore, when designing the coupling matrix, not only the stability of a pattern as a fixed point but also the size of its basin of attraction are important measures. The most classical learning rule for coupling matrices is the Hebb rule [4]. Such a rule defines the

coupling matrix $\mathbf{J}$ for P patterns $\{\xi^\mu\}_{\mu=1,\dots,P}$ with $\xi_i^\mu \in \{+1, -1\}$, as follows

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu \quad \text{if } i \neq j, \quad J_{ii} = 0 \quad (2)$$

The Hebb rule has long served as the standard form of the Hopfield model because it is biologically plausible [5] and tractable to theoretical analysis [6]. However, the Hebb rule has inherent limitations. First, its storage capacity remains limited for a network with N neurons, as the critical load is approximately $\alpha_c = P_c/N \approx 0.138$ [6]. Second, as the load increases, mixed states and spurious attractors other than the stored patterns become more numerous, and retrieval success depends strongly on the initial condition. Therefore, a central challenge in the Hopfield model is to increase storage capacity and enlarge the basin of attraction while preserving biological plausibility.

Many learning rules have been proposed to address this challenge. A representative analytical approach is the pseudo-inverse rule [7, 8]. This rule has a major advantage in that it can construct fixed points without retrieval errors, even for correlated patterns. On the other hand, this approach requires the inversion of the pattern correlation matrix, which makes it non-local and not always suitable for the sequential addition of new memories. It is also known that the basin of attraction becomes small as the load approaches the theoretical capacity. Iterative learning rules, such as unlearning or dreaming, have also been proposed [9]. These methods remove undesirable stable states by progressively weakening spurious memories. These learning rules consist of local updates, which also make them biologically plausible. However, classical dreaming can suffer from catastrophic forgetting, in which excessive iterations deteriorate even the memories the network should retain.

Recently, a lot of research activity has explored the dreaming mechanics, and generalizations of it, with the aim of better understanding how it manages to increase the network capacity [1, 10–15].

The algorithm that at present is the most effective in increasing the capacity of the network and the size of the basin of attraction is the Daydreaming algorithm proposed in Ref. [1]. Daydreaming is a learning algorithm that maintains good performance even after long learning by combining Hebbian updates that repeatedly reinforce the patterns to be stored with anti-Hebbian (i.e., unlearning or dreaming) updates that weaken spurious attractors reached from random initial conditions. Although Daydreaming has similarities to Boltzmann machine learning in that it combines positive and negative updates, it differs fundamentally in that the states to be unlearned are sampled via zero-temperature dynamics from random initial states, rather than by equilibrium finite-temperature sampling, so that Daydreaming is intrinsically an out-of-equilibrium process. The authors of Ref. [1] showed that Daydreaming not only performs well for uncorrelated patterns, but also forms large basins of attraction for correlated patterns generated by the random features model [16, 17]. It has been further reported that Daydreaming also works effectively on more realistic data, like the MNIST dataset. These results show that Daydreaming provides a promising approach to improve the performance of the Hopfield model.

However, the existing formulation of Daydreaming has mainly focused on unbiased patterns, where each element has zero mean, or on structured data consistent with that assumption, and a formulation for biased patterns has not yet been fully established. We define biased patterns as a set of patterns whose elements satisfy $\mathbb{E}[\xi_i^\mu] = m \neq 0$. These kinds of settings naturally arise when we consider neural representations with low activity rates [18], and theory has long dealt with them as important problems [19].

Previous studies showed that a naive application of the Hopfield model to biased random patterns fails because the crosstalk noise generated by other stored patterns has a non-zero mean, which can destabilize the stored patterns even at low load [19]. They further showed that subtracting the pattern bias when constructing the coupling matrix restores the local stability of retrieval states, but is not, by itself, sufficient to ensure reliable retrieval. Therefore, for biased patterns, the essential issue is not only how to modify the coupling matrix but also how to define dynamics consistent with the prescribed mean activity. This issue becomes central when we extend Daydreaming for biased patterns. A naive extension of Daydreaming is to introduce a local field that controls the bias, in addition to a centered coupling matrix obtained from subtracting the pattern means. We believe this formulation is reasonable in itself. However, when evaluating performance on biased patterns, learning rules such as the pseudo-inverse rule require additional hyperparameter tuning to control the mean activity, making a fair comparison difficult.

In this study, we reformulate Daydreaming for biased patterns inspired by the pseudo-inverse rule. Specifically, we first revisit the centered pseudo-inverse rule for biased patterns and then clarify the corresponding retrieval dynamics and energy function that naturally derive from it.

Next, we introduce a Daydreaming update rule for the centered coupling matrix that is consistent with this framework. As a result, the *centered Daydreaming* we introduce in this paper is not an *ad hoc* modification for biased patterns, but a learning rule defined on the same centered representation as the pseudo-inverse rule. Moreover, the conventional Daydreaming is one of the special forms of centered Daydreaming proposed in this study.

In the remainder of this paper, we first describe the setting of biased patterns. We then formulate centered dynamics from the pseudo-inverse rule for biased patterns. On this basis, we propose the centered Daydreaming algorithm. Finally, through numerical experiments on biased patterns, we compare centered Daydreaming with the centered pseudo-inverse rule in terms of retrieval maps, basin sizes, and spectral properties. We further investigate how the eigenvalue spectrum affects retrieval performance using modified coupling matrices.

2 Problem Setting and Formulation

In this section, we first describe the setting considered in this study and then formulate centered Daydreaming for biased patterns.

2.1 Problem Setting and Centered Representation

We consider a Hopfield model composed of N binary neurons, and we denote its state by $\mathbf{s} \in \{+1, -1\}^N$. The stored patterns are $\{\xi^\mu\}_{\mu=1,\dots,P}$ where each element $\xi_i^\mu \in \{+1, -1\}$. In the unbiased patterns, each element has a zero mean. In this study, we consider patterns that follow the distribution shown below, which includes unbiased patterns as a special case.

$$\xi_i^\mu = \begin{cases} +1 & \text{with probability } p_1, \\ -1 & \text{with probability } 1 - p_1. \end{cases} \quad (3)$$

When $p_1 \neq 0.5$, the mean of each element is generally non-zero, and the set of patterns has a uniform bias. We define the mean of each element across the pattern set as

$$m_i \equiv \frac{1}{P} \sum_{\mu=1}^P \xi_i^\mu. \quad (4)$$

Here, $\mathbf{m} = (m_1, \dots, m_N)^T$ denotes the neuron-wise mean. When we deal with biased patterns, it is natural to work not with the original pattern ξ^μ itself, but with a centered representation obtained by subtracting the mean. This representation plays an important role in defining both the pseudo-inverse rule and Daydreaming consistently for biased patterns.

$$\tilde{\xi}^\mu \equiv \xi^\mu - \mathbf{m} \quad (5)$$

2.2 Pseudo-Inverse Rule for Biased Patterns

For unbiased patterns, the pseudo-inverse rule defines the coupling matrix by using the pattern matrix $\Xi = (\xi^1, \dots, \xi^P) \in \{+1, -1\}^{N \times P}$ and the correlation matrix $\mathbf{C} = \frac{1}{N} \Xi^T \Xi \in \mathbb{R}^{P \times P}$ as follows

$$\mathbf{J} = \frac{1}{N} \Xi \mathbf{C}^{-1} \Xi^T. \quad (6)$$

Now, the relation $\Xi = \mathbf{J} \Xi$, together with $\xi_i^\mu \in \{+1, -1\}$, guarantees that each pattern ξ^μ satisfies the fixed-point condition of the asynchronous retrieval dynamics. For biased patterns, we use the following centered representation matrix rather than the original pattern matrix

$$\tilde{\Xi} \equiv \Xi - \mathbf{m} \mathbf{1}^T. \quad (7)$$

Here, $\mathbf{1}$ denotes the P -dimensional vector whose elements are all equal to 1. Since the centering is performed with the empirical mean over the stored patterns, the centered representation matrix satisfies $\tilde{\Xi} \mathbf{1} = \mathbf{0}$. Therefore, the corresponding centered correlation matrix is generally singular. We denote the effective rank of the centered representation matrix by

$$r \equiv \text{rank}(\tilde{\Xi}) = \text{rank}(\tilde{\mathbf{C}}). \quad (8)$$

The empirical centering condition $\tilde{\Xi} \mathbf{1} = \mathbf{0}$ implies $r \leq \min(N, P - 1)$. For independently generated random patterns with $0 < p_1 < 1$ and $P - 1 \leq N$, this centering condition is typically the only linear dependence among the columns of $\tilde{\Xi}$, and hence the typical rank is $r = P - 1$. We thus

define the centered correlation matrix $\tilde{\mathbf{C}}$ and the centered pseudo-inverse coupling matrix $\tilde{\mathbf{J}}$ for biased patterns as

$$\tilde{\mathbf{C}} \equiv \frac{1}{N} \tilde{\mathbf{\Xi}}^T \tilde{\mathbf{\Xi}}, \quad (9)$$

$$\tilde{\mathbf{J}} \equiv \frac{1}{N} \tilde{\mathbf{\Xi}} \tilde{\mathbf{C}}^+ \tilde{\mathbf{\Xi}}^T, \quad (10)$$

where $\tilde{\mathbf{C}}^+$ denotes the Moore-Penrose pseudoinverse of $\tilde{\mathbf{C}}$. Although $\tilde{\mathbf{C}}^+ \tilde{\mathbf{C}}$ is a projector rather than the identity matrix, the following relation still holds:

$$\tilde{\mathbf{\Xi}} = \tilde{\mathbf{J}} \tilde{\mathbf{\Xi}}. \quad (11)$$

Eq. (11) implies that the following equation is satisfied for each pattern ξ^μ

$$\xi^\mu - \mathbf{m} = \tilde{\mathbf{J}}(\xi^\mu - \mathbf{m}) \iff \xi^\mu = \tilde{\mathbf{J}}(\xi^\mu - \mathbf{m}) + \mathbf{m}. \quad (12)$$

The important point here is that $\tilde{\mathbf{J}}$ contains information not on the pattern itself but on its deviation from the mean.

2.3 Retrieval Dynamics and the Corresponding Energy Function

Motivated by Eq. (12), and following the standard Hopfield convention of excluding self-couplings, we define the retrieval dynamics for biased patterns as

$$s_i^{(t+1)} = \text{sign} \left(\sum_{j \neq i} \tilde{J}_{ij} (s_j^{(t)} - m_j) + m_i \right). \quad (13)$$

This dynamics extends the retrieval dynamics for unbiased patterns to the centered representation obtained by subtracting the mean $\mathbf{m}$ from the state $\mathbf{s}$, and then considers the shift from the mean afterward. Indeed, when $m_i = 0$ holds for all i , this equation reduces to the standard update rule Eq. (1).

From the update rule in Eq. (13), we derive the corresponding energy function

$$\mathcal{H}(\mathbf{s}) = -\frac{1}{2} \sum_{i \neq j} \tilde{J}_{ij} (s_i - m_i)(s_j - m_j) - \sum_{i=1}^N m_i s_i. \quad (14)$$

The first term represents the pairwise interaction for the centered representation $\mathbf{s} - \mathbf{m}$, and the second term provides an effective field induced by the bias.

This perspective is important because it allows us to treat both the pseudo-inverse rule and Daydreaming on the same basis. If we were to apply the pseudo-inverse rule to biased patterns in the same way as Ref. [19], we should first construct the centered coupling matrix and then we should tune the hyperparameter g that controls the mean activity, which makes a fair comparison with Daydreaming difficult. To avoid this complication, we formulate the problem by specifying only the centered coupling matrix $\tilde{\mathbf{J}}$ for biased patterns.

2.4 Centered Daydreaming

Daydreaming is an iterative learning rule that simultaneously reinforces the patterns to be stored and unlearns the fixed points reached from random initial states. In Daydreaming, at each step u , we randomly select a stored pattern $\xi^{\mu(u)}$, run the retrieval dynamics from a random initial state and acquire a fixed point $\sigma^{(u)}$, and then update the coupling matrix $\mathbf{J}$ as follows

$$J_{ij}^{(u+1)} = J_{ij}^{(u)} + \frac{1}{\tau N} \left(\xi_i^{\mu(u)} \xi_j^{\mu(u)} - \sigma_i^{(u)} \sigma_j^{(u)} \right). \quad (15)$$

The parameter τ is the inverse learning rate, and can be fixed without any particular effort [1].

The update rule of the *centered Daydreaming* we propose in this study is given as follows

$$\tilde{J}_{ij}^{(u+1)} = \tilde{J}_{ij}^{(u)} + \frac{1}{\tau N} \left[\left(\xi_i^{\mu(u)} - m_i \right) \left(\xi_j^{\mu(u)} - m_j \right) - \left(\sigma_i^{(u)} - m_i \right) \left(\sigma_j^{(u)} - m_j \right) \right]. \quad (16)$$

Here, $\sigma^{(u)}$ is the fixed point obtained by running the centered retrieval dynamics introduced in Eq. (13) from a random initial state until convergence. Algorithm 1 presents the pseudocode for the complete numerical procedure.

Thus, the centered Daydreaming proposed in this study is not an empirical variant obtained by just adding a bias-correction term, but a learning rule derived from the centered pseudo-inverse rule. In the numerical experiments that follow, we apply this centered Daydreaming to biased patterns and compare it with the centered pseudo-inverse rule to show how bias affects retrieval performance, the basin of attraction, and the spectral structure of the coupling matrix.

Algorithm 1 centered Daydreaming learning algorithm

Require: patterns $\{\xi^\mu\}_{\mu=1}^P$

- 1: $m_i \leftarrow \frac{1}{P} \sum_{\mu=1}^P \xi_i^\mu$ ▷ Calculate the bias of each element
- 2: $\tilde{J}_{ij} \leftarrow \frac{1}{N} \sum_{\mu=1}^P (\xi_i^\mu - m_i)(\xi_j^\mu - m_j)$ ▷ Initialization with the centered Hebb rule
- 3: $\tilde{J}_{ii} \leftarrow 0$
- 4: **for** $t = 1, \dots, E$ **do** ▷ Do E epochs
- 5: **for** $u = 1, \dots, N$ **do** ▷ Do N steps in each epoch
- 6: $\mu \leftarrow \text{Unif}(\{1, \dots, P\})$ ▷ Pick a pattern at random
- 7: $\sigma_i \leftarrow \text{Unif}(\{+1, -1\}) \forall i$ ▷ Initialize state at random
- 8: **while** not converged **do**
- 9: $\sigma_i \leftarrow \text{sign} \left(\sum_{j \neq i} \tilde{J}_{ij}(\sigma_j - m_j) + m_i \right)$ ▷ Run the retrieval dynamics
- 10: **end while**
- 11: $\tilde{J}_{ij} \leftarrow \tilde{J}_{ij} + \frac{1}{\tau N} [(\xi_i^\mu - m_i)(\xi_j^\mu - m_j) - (\sigma_i - m_i)(\sigma_j - m_j)]$
- 12: ▷ Update the centered coupling matrix
- 13: $\tilde{J}_{ii} \leftarrow 0$
- 14: **end for**
- 15: **end for**

3 Experimental Setup

In this section, we describe the numerical experiment setup to evaluate the performance of the centered pseudo-inverse rule and centered Daydreaming for biased patterns.

3.1 Pattern Generation

In this study, we use biased random patterns, where each element is generated independently, as defined in Eq. (3). The case $p_1 = 0.5$ corresponds to unbiased patterns, while $p_1 > 0.5$ means biased patterns in which the positive element is dominant. In this study, we systematically vary p_1 as a control parameter to examine how the strength of the bias affects retrieval performance.

3.2 Evaluation Metrics

We evaluate the performance of each learning rule by using the magnetization m^μ , which is the overlap between a stored pattern ξ^μ and the state $\mathbf{s}$

$$m^\mu = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu s_i. \quad (17)$$

This quantity measures how the state matches the stored pattern. When $m^\mu = 1$, the state $\mathbf{s}$ matches the pattern ξ^μ exactly. By contrast, when $m^\mu = 0$, the state $\mathbf{s}$ is orthogonal to the pattern ξ^μ , which means that it contains no information about that pattern.

An important measure is the size of the basin of attraction. Specifically, for each pattern ξ^μ , we prepare an initial state $\mathbf{s}_{\text{init}}$ with an initial overlap m_{init}^μ and we run the centered retrieval dynamics shown in Eq. (13) until it converges to a fixed point. We then compute the final overlap m_{final}^μ from the converged state $\mathbf{s}_{\text{final}}$. We can interpret a learning rule as having a larger basin of attraction when it exhibits a wider plateau in m_{init}^μ such that $m_{\text{final}}^\mu \approx 1$. Many studies have used this measure, and we also use it as a key indicator for comparison in this study.

3.3 Experimental Details

Throughout all numerical experiments, including the retrieval map and spectral analyses, we followed the standard Hopfield-network convention of excluding self-couplings by setting the diagonal elements of coupling matrices to zero, $\tilde{J}_{ii} = 0$. Since the exact projector relation of the centered pseudo-inverse matrix Eq. (11) is defined for the matrix before removing diagonal terms, we examine the effect of retaining self-couplings in Appendix B. As a representative example, we fixed $N = 500$ and $\alpha = 0.4$, varying the bias parameter p_1 to study the effect of the bias. We compared the centered pseudo-inverse rule with centered Daydreaming. For centered Daydreaming, we fixed the inverse learning rate at $\tau = 256$, a value shown to be sufficiently large for unbiased patterns in Ref. [1], and examined how the performance changed over epochs. We initialized $\tilde{\mathbf{J}}$ in centered Daydreaming by using the centered Hebb rule. To draw the retrieval map, we generated five initial states $\mathbf{s}_{\text{init}}$ with overlap m_{init}^μ for every pattern ξ^μ . We performed numerical experiments on 20 instances for each value of the bias parameter p_1 .

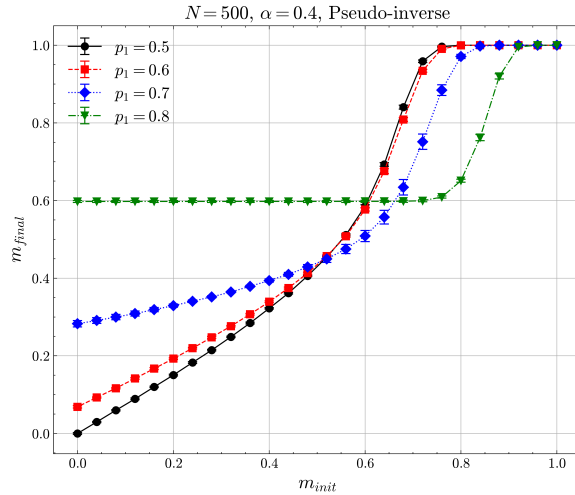


Figure 1: Retrieval maps obtained using the centered pseudo-inverse rule: each curve corresponds to a different bias strength.

4 Results

In this section, we present numerical results for the centered pseudo-inverse rule and the centered Daydreaming applied to biased patterns. We first present the retrieval maps and the basins of attraction for each learning rule. Next, we compare the spectral structure of the coupling matrix $\tilde{\mathbf{J}}$. Finally, we show retrieval maps of modified coupling matrices to probe how the broadness of the leading eigenvalue spectrum affects retrieval performance.

4.1 Retrieval Maps and Basin of Attraction Size

First, we show the retrieval maps for the centered pseudo-inverse rule in Figure 1. For $p_1 = 0.5$, which corresponds to unbiased patterns, a clear plateau with $m_{\text{final}} \approx 1$ appears in the region with relatively large initial magnetization, $m_{\text{init}} \gtrsim 0.75$. It confirms that the pseudo-inverse rule successfully retrieves the stored patterns. However, as we increase the bias, the shape of the retrieval map changes. At $p_1 = 0.6$, the map still exhibits a similar plateau, but at $p_1 = 0.7$, the network requires a larger initial overlap to reach $m_{\text{final}} \approx 1$. At $p_1 = 0.8$, the plateau with $m_{\text{final}} \approx 1$ has shrunk a lot, $m_{\text{init}} \gtrsim 0.9$, making pattern retrieval a very inefficient and practically impossible task even for a moderate level of noise in the data.

One may be surprised that, in this unfavorable case ($p_1 = 0.8$), the retrieval map in Figure 1 shows a rather large value $m_{\text{final}} \approx 0.6$ at very low values of m_{init} . However, this value is what one gets if the matrix $\tilde{\mathbf{J}}$ provides no information at all and centered retrieval dynamics in Eq. (13) bring the system to a state where every neuron is given by $s_i = \text{sign}(m) = 1$. Indeed, in this case, one has $m_{\text{final}} = 2p_1 - 1$, but this value brings no information on the pattern to be retrieved.

Next, we present the retrieval maps for the centered Daydreaming in Figure 2. At $p_1 = 0.5$ (upper left panel), the retrieval map improves monotonically with the number of learning epochs, and after a sufficiently long learning, a plateau with $m_{\text{final}} \approx 1$ emerges for $m_{\text{init}} \gtrsim 0.7$, in agreement with the results of Ref. [1]. The key observation is that the same plateau ($m_{\text{final}} \approx 1$ for $m_{\text{init}} \gtrsim 0.7$) is reached after learning by centered Daydreaming for any bias strength ($0.5 \leq p_1 \leq 0.8$) that we have studied. This means the centered Daydreaming algorithm is robust with respect to the bias in the data to be stored in a Hopfield model. This finding makes the centered Daydreaming algorithm a general and robust procedure to train a Hopfield network with realistic data. Notice that the centered Daydreaming algorithm has just one parameter, τ , which can be fixed without any particular effort (as in the original Daydreaming algorithm). So, centered Daydreaming is a ready-to-use, general and robust algorithm.

We also checked the dependence on the load, $\alpha = P/N$. The results reported in Appendix C show that centered Daydreaming retains plateaus even when α is increased, both for unbiased and biased patterns. This confirms that the robustness to bias observed at $\alpha = 0.4$ is not restricted to a single load value.

4.2 Spectral Properties of the Coupling Matrix

The eigenvalue distributions of $\tilde{\mathbf{J}}$ obtained from the centered pseudo-inverse rule and the centered Daydreaming differ for every value of p_1 as shown in Figure 3.

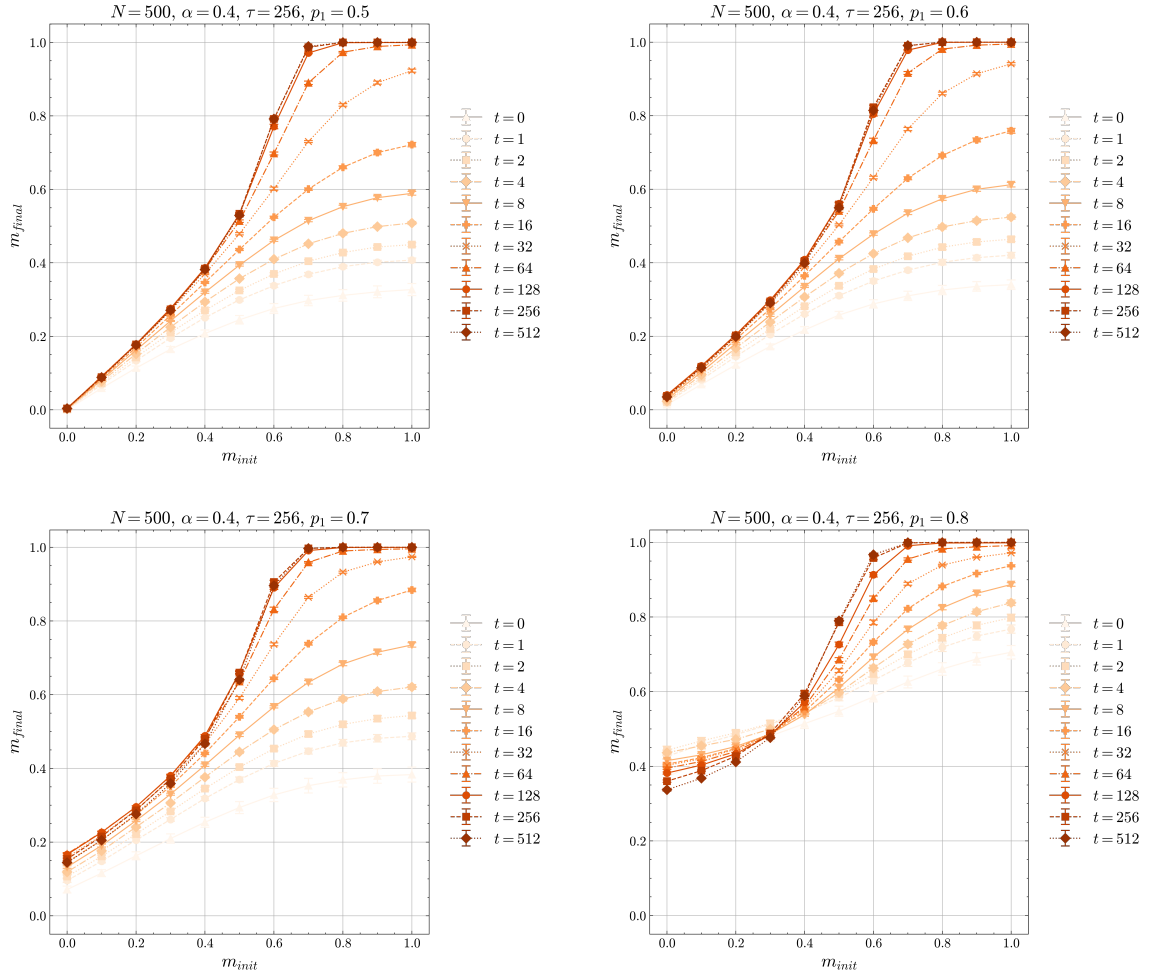


Figure 2: Retrieval maps obtained by training the coupling matrix by the centered Daydreaming algorithm: each panel differs in the value of bias parameter p_1 .

The spectrum of the centered pseudo-inverse rule is concentrated in two spectral sectors: the leading $r = P - 1$ modes are associated with the subspace spanned by the centered representations of the P patterns and are responsible for bringing the retrieval dynamics towards one of the patterns; the remaining $N - r$ modes correspond to the null space (recall that the coupling matrix has rank $r = P - 1$ because of the centering condition). The double-peak shape of the spectrum is preserved for all p_1 values.

The spectrum of the matrix learned via the centered Daydreaming algorithm is different: the negative peak (corresponding to the null space) is still very narrow, but the distribution of positive eigenvalues is much broader, with a maximum eigenvalue that is more than twice that of the pseudo-inverse and a tail extending down to the origin. What we observe in Figure 3 for Daydreaming is consistent with the picture reported in previous studies on dreaming-based learning, where the learning process selectively reshapes the upper part of the spectrum of the coupling matrix [14].

Given that the coupling matrix obtained by Daydreaming shows better statistical properties in terms of basins of attraction and, consequently, in terms of maximum capacity, it is worth trying to relate this better performance to the broadness of the spectrum shown in Figure 3. In both cases, pseudo-inverse and Daydreaming, the leading r eigenvectors are expected to span the subspace defined by the centered representations of the P patterns. Thus, the application of the matrix $\tilde{\mathbf{J}}$ to the centered state vector $\mathbf{s} - \mathbf{m}$ during the retrieval dynamics in Eq. (13) tends to project the centered state into this subspace. What differs between the two rules is that, in the pseudo-inverse case, all these leading modes have approximately the same weight, while in the Daydreaming case, the weights used to project the centered state vector are very diverse. This induces a sort of hierarchy among eigenvectors in the Daydreaming case, such that the centered state vector is first projected towards the eigenvectors with the largest eigenvalues, almost ignoring the other eigenvectors; these are considered in later stages of the dynamics. This observation suggests that

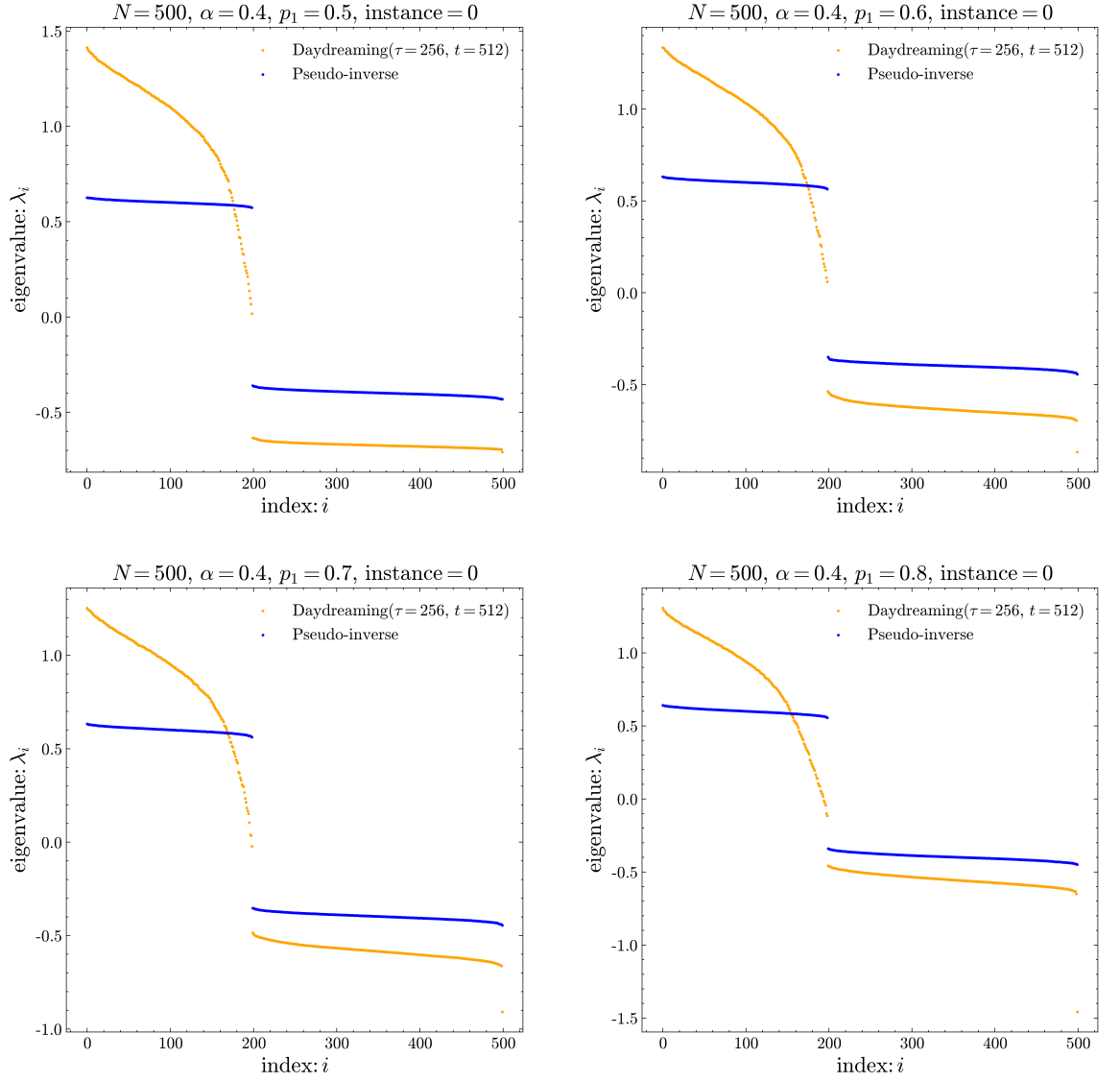


Figure 3: Eigenvalue distributions for the pseudo-inverse rule and Daydreaming. For Daydreaming, we plot the eigenvalue distribution at epoch $t = 512$ for a representative instance.

the enlarged basin of attraction is related not only to the subspace spanned by the leading r eigenvectors, but also to the non-uniform weights assigned to these modes by the first r eigenvalues. To evaluate this point quantitatively, in the next subsection, we construct modified coupling matrices that keep the eigenvectors fixed while controlling the broadness of the leading r eigenvalues.

4.3 Retrieval Maps of Modified Coupling Matrices

We probe how the broadness of the leading $r = P - 1$ modes discussed in the previous subsection affects retrieval performance. Specifically, we construct modified coupling matrices that interpolate between the broad leading spectrum generated by Daydreaming and a pseudo-inverse-like spectrum. We then use these modified coupling matrices to draw retrieval maps and quantify the relationship between spectral broadness and retrieval performance.

First, we perform a spectral decomposition of the coupling matrix learned by Daydreaming and obtain its eigenvalues and eigenvectors. We use the Daydreaming coupling matrix at $t = 512$ as the original coupling matrix to be spectrally decomposed. We denote this matrix by $\tilde{\mathbf{J}}_{\text{Daydreaming}}$ and write its spectral decomposition as

$$\tilde{\mathbf{J}}_{\text{Daydreaming}} = \sum_{i=1}^N \lambda_i \mathbf{v}_i \mathbf{v}_i^T, \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N. \quad (18)$$

Here, we assume that the eigenvalues are sorted in descending order. Next, for the leading r

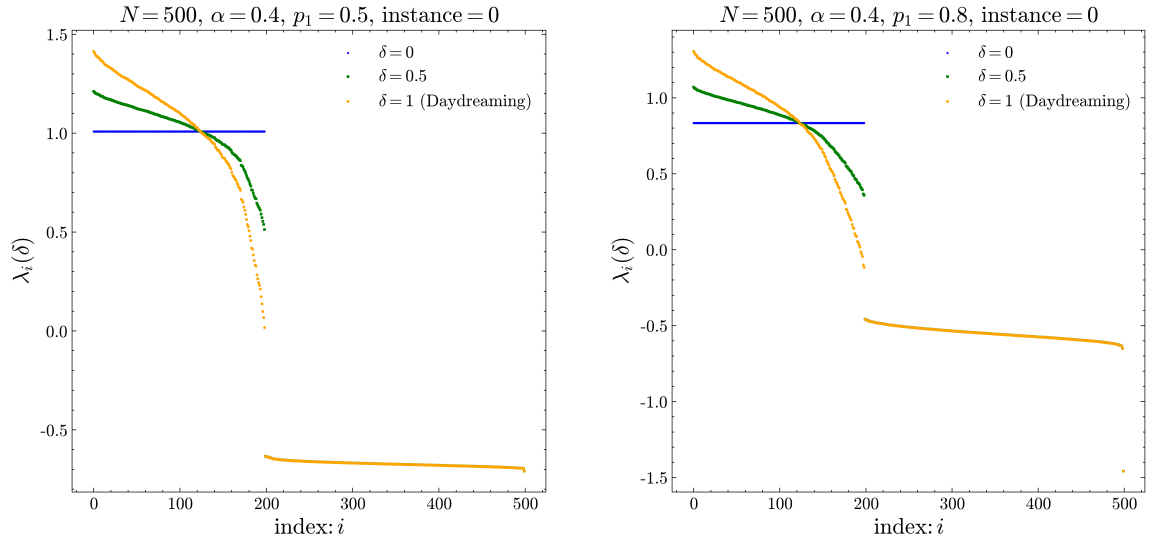


Figure 4: Designed eigenvalue spectra parametrized by δ and used to construct the modified coupling matrices, shown for a representative instance: each panel differs in the value of bias parameter p_1 .

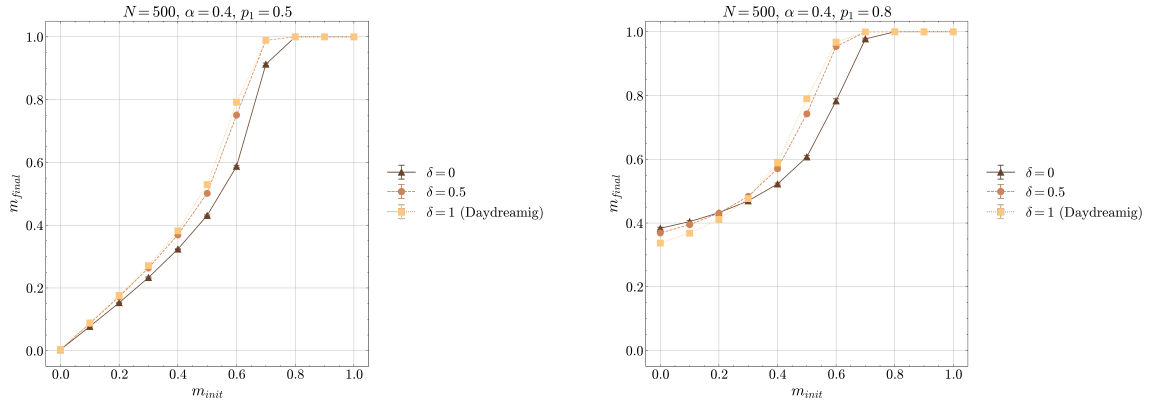


Figure 5: Retrieval maps of the modified coupling matrices for $p_1 = 0.5$ and $p_1 = 0.8$. The parameter δ controls the broadness of the leading eigenvalue spectrum, with $\delta = 1$ corresponding to the original Daydreaming spectrum and $\delta = 0$ to the pseudo-inverse-like spectrum.

eigenvalues associated with the subspace spanned by the centered representations of the P patterns, we introduce a transformation parameterized by $\delta \in [0, 1]$. Defining the mean value $\bar{\lambda} = \sum_{\mu=1}^r \lambda_{\mu}/r$, we have

$$\lambda_{\mu}(\delta) = (1 - \delta)\bar{\lambda} + \delta\lambda_{\mu}, \quad \mu = 1, \dots, r. \quad (19)$$

The case $\delta = 1$ corresponds to the original Daydreaming spectrum, whereas $\delta = 0$ corresponds to a pseudo-inverse-like setting in which all leading r modes have equal weights. The remaining $N - r$ eigenvalues and all eigenvectors are kept unchanged. We show specific examples of modified spectra for $p_1 = 0.5$ and $p_1 = 0.8$ in Figure 4. Using the spectrum characterized by δ , we construct the modified coupling matrix as

$$\tilde{\mathbf{J}}_{\text{modified}}(\delta) = \sum_{\mu=1}^r \lambda_{\mu}(\delta) \mathbf{v}_{\mu} \mathbf{v}_{\mu}^T + \sum_{\nu=r+1}^N \lambda_{\nu} \mathbf{v}_{\nu} \mathbf{v}_{\nu}^T. \quad (20)$$

We examine the retrieval maps obtained by varying δ in the modified coupling matrix¹. For each of $p_1 = 0.5$ and $p_1 = 0.8$, we construct the modified coupling matrices for all 20 instances and draw the corresponding retrieval maps. We show the results in Figure 5. As a result, for both $p_1 = 0.5$ and $p_1 = 0.8$, we observed that the plateau region with $m_{\text{final}} \approx 1$ remains unchanged

¹Before using the modified coupling matrices $\tilde{\mathbf{J}}_{\text{modified}}(\delta)$ in the retrieval dynamics, we apply the same zero-diagonal convention as in the other experiments.

going from $\delta = 1$ to $\delta = 0.5$, while it clearly shrinks at $\delta = 0$. Thus, we conclude that the spectral broadness (keeping eigenvectors fixed) certainly contributes to the enlargement of the basins of attraction in the retrieval dynamics.

5 Conclusions

In this study, we have introduced a *centered* version of the Daydreaming algorithm that can work effectively in the tested biased settings. Increasing the bias in the data, the coupling matrix trained by the centered Daydreaming algorithm presents larger basins of attraction with respect to the centered pseudo-inverse and to the original Daydreaming algorithm. Training by centered Daydreaming, the size of the basin of attraction stays almost constant in a broad range of biases ($0.5 \leq p_1 \leq 0.8$). These results make the centered Daydreaming algorithm of practical use in realistic situations where data can be structured and biased.

We have also shown that the coupling matrix learned by centered Daydreaming has a broader eigenvalue spectrum than that obtained with the centered pseudo-inverse rule. Experiments with modified coupling matrices, in which the leading r eigenvalues were averaged out toward their mean while the eigenvectors were kept fixed, provide evidence that the broadness of the weights/eigenvalues of the leading modes is associated with enlarged basins of attraction. They show that a pseudo-inverse-like leading spectrum shrinks the retrieval plateau, indicating that the non-uniform weighting of the leading r modes affects retrieval performance and contributes to enlarging the basin of attraction. This finding suggests that centered Daydreaming shapes a richer, possibly hierarchical, energy landscape that helps retrieval dynamics.

Supplementary experiments (shown in the Appendices) further support the effectiveness of centered Daydreaming. The load-dependence shows that the robustness to bias is not restricted to the representative load. The biased Random-Features Hopfield Model experiment provides an additional example in which centered Daydreaming forms retrieval plateaus for structured biased patterns. However, the slight degradation after long learning epochs and the behavior of feature magnetization indicate that the mechanism in correlated biased data is not yet fully understood.

Future work will include a more precise analysis of how the breadth of the leading eigenvalue spectrum affects retrieval performance, and a deeper understanding of centered Daydreaming for more structured, biased patterns, such as biased Random-Features Hopfield Model across a wider range of α and α_D .

Appendix A. Original Daydreaming for Biased Patterns

For completeness, we also present in Figure 6 the results obtained by applying the original Daydreaming algorithm of Ref. [1] with biased patterns. With respect to the results obtained by the centered version and reported in Figure 2, when the bias parameter p_1 becomes large, we observe that (i) the plateau with $m_{\text{final}} \approx 1$ is smaller and (ii) the time required by the learning process to converge is larger. For example, in the $p_1 = 0.8$ case (lower right panel), at time $t = 512$ the learning process has still not reached convergence, and indeed the plateau with $m_{\text{final}} \approx 1$ is still not formed. Clearly, the centered version of the Daydreaming algorithm introduced in the present work has much better performance than the original Daydreaming algorithm.

Appendix B. Effect of self-couplings in the centered pseudo-inverse rule

In the main text, we evaluated the retrieval performance of the centered pseudo-inverse rule, excluding the effect of self-couplings. However, the exact projector relation of the centered pseudo-inverse matrix is defined for the matrix before the diagonal interactions are removed. Therefore, to clarify the role of the diagonal interactions, we also evaluate the centered pseudo-inverse dynamics with self-couplings included:

$$s_i^{(t+1)} = \text{sign} \left(\sum_{j=1}^N \tilde{J}_{ij} \left(s_j^{(t)} - m_j \right) + m_i \right). \quad (21)$$

This dynamics preserves the exact fixed-point property of the full centered pseudo-inverse matrix. We show the retrieval map at $\alpha = 0.4$ in Figure 7. Despite the exact fixed-point property, retaining self-couplings does not improve the retrieval performance drastically. This result indicates that exact stability of the stored patterns does not necessarily imply a large basin of attraction. We can understand this behavior from the form of the self-coupled local field. Including self-couplings, the retrieval dynamics differ from the one in Eq. (13) by the additional term $\tilde{J}_{ii} \left(s_i^{(t)} - m_i \right)$. This term depends on the current neuron value and therefore acts as an inertia-like contribution because

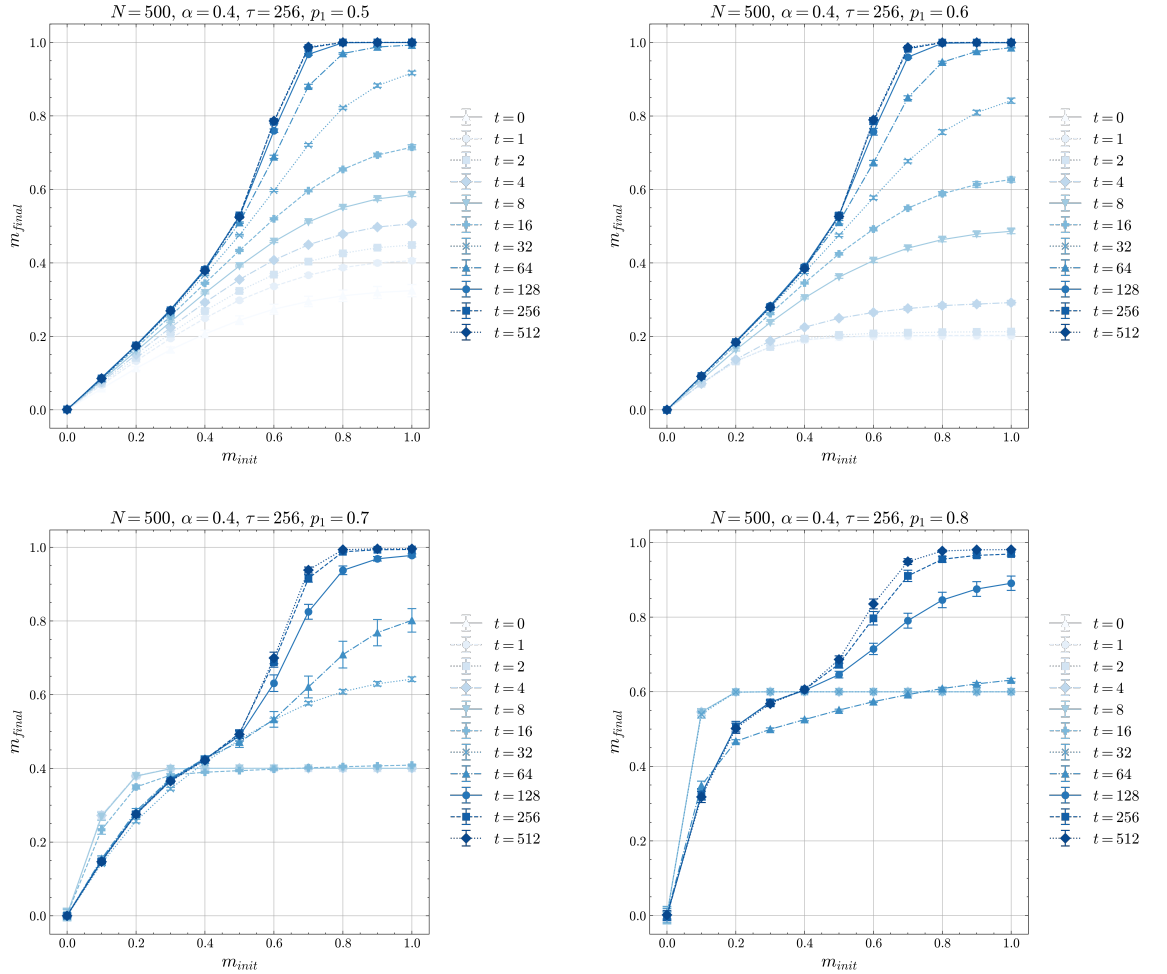


Figure 6: Retrieval map for the original Daydreaming algorithm: panels differ by the value of bias parameter p_1 .

typically $\tilde{J}_{ii} > 0$. It stabilizes neurons that are already aligned with a stored pattern, but it can also hold erroneous neurons in noisy initial states. Consequently, restoring the exact fixed-point property by retaining self-couplings does not necessarily enhance the dynamics' error-correction ability. These results suggest that the superiority of centered Daydreaming over the centered pseudo-inverse rule cannot be attributed simply to the removal of self-couplings from the centered pseudo-inverse matrix.

Appendix C. Load dependence with respect to α

In the main text, we fixed the load at $\alpha = P/N = 0.4$ to clearly compare the effect of the bias parameter p_1 . In this Appendix, we present additional experiments in which α is varied to examine whether the robustness of centered Daydreaming against bias observed in the main text is specific to this particular load value.

The experimental setup is the same as in the main text. We set $N = 500$ and used $P = \alpha N$ stored patterns. For each set of patterns, we computed the neuron-wise mean m_i and learned the coupling matrix using centered Daydreaming. We fixed the inverse learning rate to $\tau = 256$, as in the main experiments. We considered $\alpha = \{0.05, 0.2, 0.6, 1.0\}$, and examined both the unbiased case with $p_1 = 0.5$ and the strongly biased case with $p_1 = 0.8$. We set the number of learning epochs to $t = 512$ for $\alpha = 0.05, 0.2$, and 0.6 . For $\alpha = 1.0$, we needed a longer learning time $t = 8192$ because the convergence of learning becomes slower. We performed numerical experiments on 20 instances for each setting.

We show the retrieval maps in Figure 8. In the case with $p_1 = 0.5$, when α is small, the system reaches the plateau with $m_{final} \approx 1$ even from small initial magnetizations. As α increases, the initial magnetization required for the retrieval to appear becomes larger. This behavior reflects the shrinkage of the basin of attraction as the number of stored patterns increases. For the case with $p_1 = 0.8$, a similar load dependence is observed. For biased patterns, the non-zero value of m_{final} in

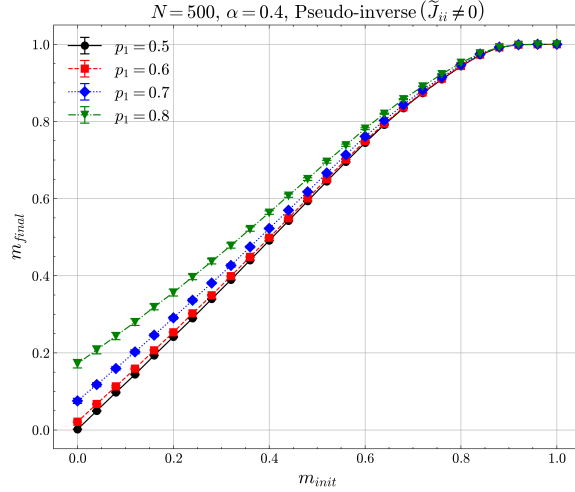


Figure 7: Retrieval maps obtained using the centered pseudo-inverse rule with self-couplings $\tilde{J}_{ii}$: each curve corresponds to a different bias strength.

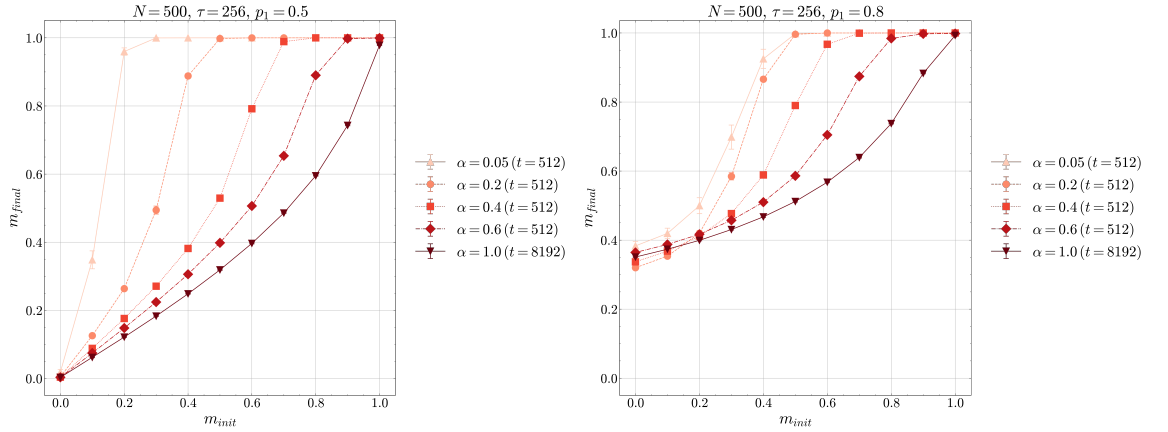


Figure 8: Load dependence of the retrieval maps obtained by centered Daydreaming. The left and right panels show the results for $p_1 = 0.5$ and $p_1 = 0.8$, respectively.

the low initial magnetization region includes the overlap induced by the bias. Therefore, it cannot be interpreted as a correct retrieval. On the other hand, the right panel of Figure 8 shows that, even in the biased case, centered Daydreaming can retrieve stored patterns from finite initial overlaps. In particular, for $\alpha < 1$, $m_{\text{final}} \approx 1$ is achieved in this retrieval region, indicating that the high-capacity behavior of original Daydreaming for unbiased patterns near $\alpha_c \approx 1$ is preserved for biased patterns. These results support the conclusion that the robustness of centered Daydreaming against bias, as shown in the main text, is not specific to the single load value $\alpha = 0.4$.

Appendix D. Biased Random-Features Hopfield Model

In the main text, we focused on biased random patterns whose elements are generated independently. However, actual memory patterns are not necessarily independent random patterns, and they may have correlations induced by underlying low-dimensional features. In this Appendix, we present supplementary experiments that extend the Random-Features Hopfield Model (RFHM) [17] to the biased case to examine whether centered Daydreaming also works for correlated biased patterns.

In the standard RFHM, structured stored patterns are generated from feature vectors and random coefficients through a linear projection and component-wise sign nonlinearity. We append a special feature vector whose components are all equal to +1 to introduce a global bias. Specifically, we set $f_{i1} = 1$ for all i and $c'_1 = \sqrt{D} \mathcal{B}$. For $k = 2, \dots, D+1$, the feature vectors have i.i.d. uniform binary entries $f_{ik} \in \{+1, -1\}$, and the coefficients c'_k are i.i.d. standard Gaussian

variables. Using these feature vectors and coefficients, the stored patterns are generated as

$$\xi_i^\nu = \text{sign} \left(\frac{1}{\sqrt{D}} \sum_{k=1}^{D+1} c_k^\nu f_{ik} \right) = \text{sign} \left(\mathcal{B} + \frac{1}{\sqrt{D}} \sum_{k=2}^{D+1} c_k^\nu f_{ik} \right). \quad (22)$$

Here, $\mathcal{B}$ is a parameter that controls the strength of the global bias. In particular, when $\mathcal{B} = 0$, this model reduces to the standard RFHM. The contribution from the non-bias features can be written as

$$\eta_i^\nu = \frac{1}{\sqrt{D}} \sum_{k=2}^{D+1} c_k^\nu f_{ik} \sim \mathcal{N}(0, 1). \quad (23)$$

The probability that each component of patterns takes the value +1 is given by

$$p_1 = \Pr(\xi_i^\nu = +1) = \Pr(\mathcal{B} + \eta_i^\nu > 0). \quad (24)$$

Therefore,

$$p_1 = \int_{-\mathcal{B}}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) = \frac{1}{2} \left[1 + \text{erf}\left(\frac{\mathcal{B}}{\sqrt{2}}\right) \right], \quad (25)$$

$$\mathcal{B} = \sqrt{2} \text{erf}^{-1}(2p_1 - 1). \quad (26)$$

In the experiments, we determine $\mathcal{B}$ from the above relation for a given p_1 . After learning, we evaluate the retrieval performance by the magnetization with respect to the stored patterns. We also measure the overlap between the state $\mathbf{s}$ and the feature vector $\mathbf{f}$ to examine the feature structure of the RFHM.

$$m^\nu = \frac{1}{N} \sum_{i=1}^N \xi_i^\nu s_i \quad (27)$$

$$\mu^k = \frac{1}{N} \sum_{i=1}^N f_{ik} s_i \quad (28)$$

In particular, for the special feature $f_{i1} = 1$, the quantity μ^1 is given by

$$\mu^1 = \frac{1}{N} \sum_{i=1}^N s_i. \quad (29)$$

We call this quantity global feature magnetization.

In the experiments, we set $N = 500$, $\alpha = P/N = 0.5$, and $\alpha_D = D/N = 0.1$. We investigate both the standard RFHM with $p_1 = 0.5$ and the strongly biased case with $p_1 = 0.8$. The inverse learning rate was fixed at $\tau = 256$ as in the main text. We performed numerical experiments on 20 instances for each setting.

We first examine the case with $p_1 = 0.5$. In the case of $\mathcal{B} = 0$, there is no global bias. Therefore, this case corresponds to the conventional RFHM. We show the results in Figure 9. The left panel shows the retrieval map for pattern magnetization m^ν , while the right panel shows the results for the feature magnetization μ^k with $k \geq 2$. After learning, retrieval plateaus with $m_{\text{final}} \approx 1$ and $\mu_{\text{final}}^k \approx 1$ are formed from finite initial overlaps. These results are consistent with the previous study [1] on the RFHM and confirm that the experimental setup properly reproduces the standard RFHM.

Next, we examine the biased case with $p_1 = 0.8$. In this case, $\mathcal{B} > 0$, and the stored patterns contain a strong global bias. Therefore, even in the low initial overlap region, a non-zero pattern magnetization appears due to the bias. Figure 10 shows the results for the pattern magnetization. As shown in the figure, even in the presence of a strong global bias, centered Daydreaming forms a retrieval plateau. However, at $t = 2048$ and $t = 4096$, we observe a small degradation in the retrieval performance compared with shorter learning epochs. This result implies that retrieval performance does not necessarily improve monotonically and converge in the biased RFHM. Figure 11 shows the results for the feature magnetization. The left panel shows the global feature magnetization μ^1 corresponding to the special feature $f_{i1} = 1$ for all i . The right panel of Figure 11 shows the feature magnetization μ^k for non-bias features with $k \geq 2$. As shown in the figure, for $p_1 = 0.8$, we do not observe either feature magnetization approaching 1. These results suggest that,

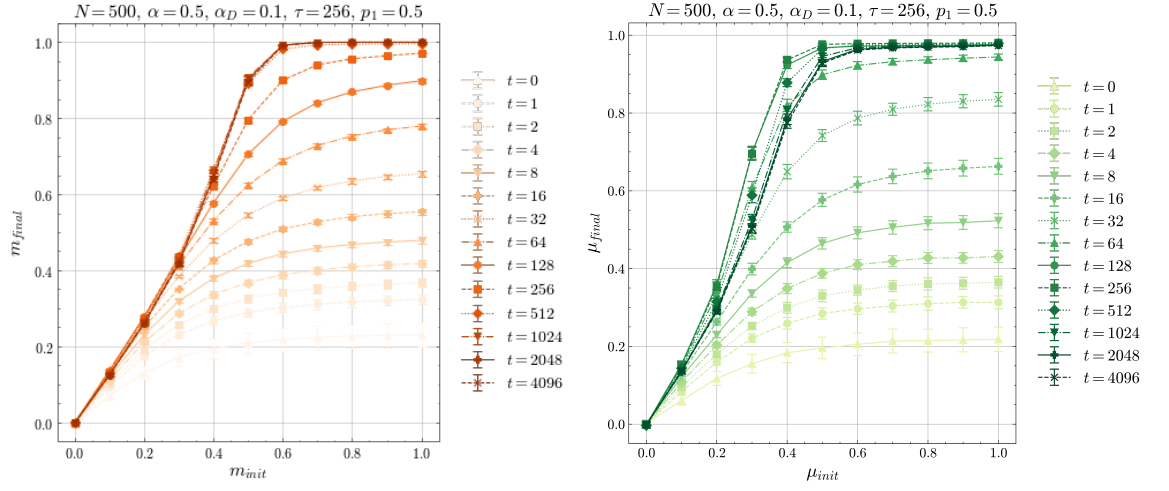


Figure 9: Retrieval maps for stored patterns and feature vectors in the standard RFHM with $p_1 = 0.5$. The left and right panels show retrieval maps for the pattern magnetization m^ν and the non-bias feature magnetization μ^k with $k \geq 2$, respectively.

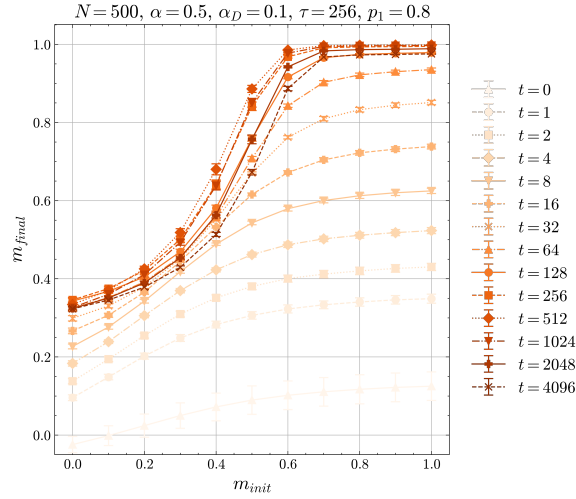


Figure 10: Retrieval map for patterns of the biased RFHM with $p_1 = 0.8$.

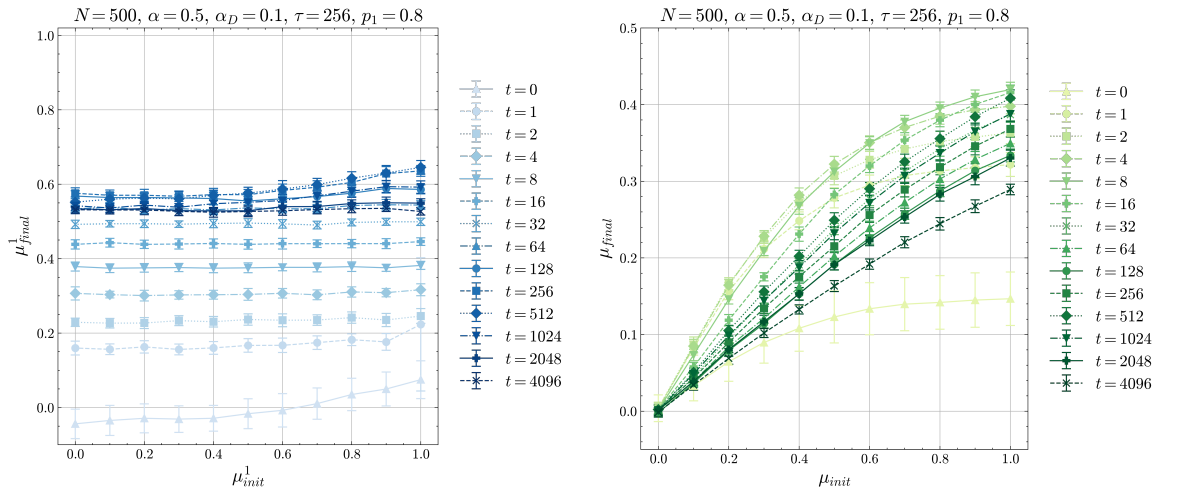


Figure 11: Retrieval maps for feature magnetization of the biased RFHM with $p_1 = 0.8$. The left panel shows the global feature magnetization μ^1 , which corresponds to the global magnetization of the state. The right panel shows the feature magnetization μ^k for non-bias features with $k \geq 2$.

in the biased RFHM, attractors are formed mainly around the stored patterns rather than spontaneously around feature vectors.

Overall, these results confirm that, under a representative setting $(\alpha, \alpha_D) = (0.5, 0.1)$, centered Daydreaming can form retrieval plateaus toward stored patterns even for correlated patterns such as the biased RFHM with a global bias. These results imply that the effectiveness shown in the main text is not limited to independently generated biased random patterns. However, understanding the mechanism behind the slight degradation in retrieval performance after long learning epochs and a more detailed verification of different settings (α, α_D) remain important topics for future work.

Acknowledgments

M.D. thanks Dario Bocchi for fruitful discussions. M.D. and M.O. were funded by programs for bridging the gap between R&D and IDEal society (Society 5.0) and Generating Economic and social value (BRIDGE) and Cross-ministerial Strategic Innovation Promotion Program (SIP) from the Cabinet Office. F.R.T. thanks Silvio Franz for raising the question that has inspired the present work and the “National Centre for HPC, Big Data and Quantum Computing”, Project CN_00000013, CUP B83C22002940006, NRRP Mission 4 Component 2 Investment 1.4, Funded by the European Union - NextGenerationEU, for the financial support.

References

- [1] Ludovica Serricchio, Dario Bocchi, Claudio Chilin, Raffaele Marino, Matteo Negri, Chiara Cammarota, and Federico Ricci-Tersenghi. Daydreaming hopfield networks and their surprising effectiveness on correlated data. *Neural Networks*, 186:107216, June 2025.
- [2] S.-I. Amari. Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Transactions on Computers*, C-21(11):1197–1206, November 1972.
- [3] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, April 1982.
- [4] Donald Olding Hebb. *The organization of behavior: A neuropsychological theory*. John Wiley and Sons, Inc., 1949.
- [5] Daniel J Amit. *Modeling brain function: The world of attractor neural networks*. Cambridge University Press, 1989.
- [6] Daniel J. Amit, Hanoach Gutfreund, and Haim Sompolinsky. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Phys. Rev. Lett.*, 55:1530–1533, Sep 1985.
- [7] L. Personnaz, I. Guyon, and G. Dreyfus. Information storage and retrieval in spin-glass like neural networks. *Journal de Physique Lettres*, 46(8):359–365, 1985.
- [8] I. Kanter and H. Sompolinsky. Associative recall of memory without errors. *Physical Review A*, 35(1):380–392, January 1987.
- [9] J. J. Hopfield, D. I. Feinstein, and R. G. Palmer. ‘unlearning’ has a stabilizing effect in collective memories. *Nature*, 304(5922):158–159, July 1983.
- [10] Elena Agliari, Francesco Alemanno, Adriano Barra, and Alberto Fachechi. Dreaming neural networks: rigorous results. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(8):083503, 2019.
- [11] Alberto Fachechi, Elena Agliari, and Adriano Barra. Dreaming neural networks: forgetting spurious memories and reinforcing pure ones. *Neural Networks*, 112:24–40, 2019.
- [12] Marco Benedetti, Enrico Ventura, Enzo Marinari, Giancarlo Ruocco, and Francesco Zamponi. Supervised perceptron learning vs unsupervised hebbian unlearning: Approaching optimal memory retrieval in hopfield-like networks. *The Journal of Chemical Physics*, 156(10), March 2022.
- [13] E. Agliari, F. Alemanno, M. Aquaro, and A. Fachechi. Regularization, early-stopping and dreaming: A hopfield-like setup to address generalization and overfitting. *Neural Networks*, 177:106389, September 2024.

- [14] Marco Benedetti, Louis Carillo, Enzo Marinari, and Marc Mézard. Eigenvector dreaming. *Journal of Statistical Mechanics: Theory and Experiment*, 2024(1):013302, January 2024.
- [15] Shuta Takeuchi, Takashi Takahashi, and Yoshiyuki Kabashima. Analysis of the hopfield model incorporating the effects of unlearning. *arXiv preprint arXiv:2602.08428*, 2026.
- [16] Sebastian Goldt, Marc Mézard, Florent Krzakala, and Lenka Zdeborová. Modeling the influence of data structure on learning in neural networks: The hidden manifold model. *Physical Review X*, 10(4), December 2020.
- [17] M. Negri, C. Lauditi, G. Perugini, C. Lucibello, and E. Malatesta. Storage and learning phase transitions in the random-features hopfield model. *Physical Review Letters*, 131(25), December 2023.
- [18] Elad Schneidman, Michael J Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.
- [19] Daniel J. Amit, Hanoeh Gutfreund, and Haim Sompolinsky. Information storage in neural networks with low levels of activity. *Physical Review A*, 35(5):2293–2303, March 1987.