Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale
2017, Vol. 71, No. 4, 299–312

1196-1961/17/$12.00 http://dx.doi.org/10.1037/cep0000115

The Brain's Representations May Be Compatible With Convolution-Based Memory Models

Kenichi Kato and Jeremy B. Caplan
University of Alberta

Convolution is a mathematical operation used in vector-models of memory that have been successful in explaining a broad range of behaviour, including memory for associations between pairs of items, an important primitive of memory upon which a broad range of everyday memory behaviour depends. However, convolution models have trouble with naturalistic item representations, which are highly auto-correlated (as one finds, e.g., with photographs), and this has cast doubt on their neural plausibility. Consequently, modellers working with convolution have used item representations composed of randomly drawn values, but introducing so-called noise-like representation raises the question how those random-like values might relate to actual item properties. We propose that a compromise solution to this problem may already exist. It has also long been known that the brain tends to reduce auto-correlations in its inputs. For example, centre-surround cells in the retina approximate a Difference-of-Gaussians (DoG) transform. This enhances edges, but also turns natural images into images that are closer to being statistically like white noise. We show the DoG-transformed images, although not optimal compared to noise-like representations, survive the convolution model better than naturalistic images. This is a proof-of-principle that the pervasive tendency of the brain to reduce auto-correlations may result in representations of information that are already adequately compatible with convolution, supporting the neural plausibility of convolution-based association-memory.

Keywords: convolution, association-memory, cued recall, representations, mathematical models

A major goal in memory research, spanning psychology, neuroscience and artificial intelligence, has been to model memory for associations (e.g., CAT–DOG, BRIDGE–LAMPPOST). Distributed memory models that have been tested on human memory behaviour typically assume that items are represented as vectors, where each dimension of the vector stands for the value of a feature of the item, although those features are usually not specified, but conceptualised as abstract. Indeed, modellers have typically made no attempt to derive item representations from real-world features, and may even have implicitly assumed that none exists.

Such models have used two major mathematical vector operations (or their close relatives): matrix outer-product (e.g., Anderson, 1970; Humphreys, Bain, & Pike, 1989; Pike, 1984; Rumelhart, Hinton, & Williams, 1986) and convolution (e.g., Longuet-Higgins, 1968; Metcalfe Eich, 1982; Murdock, 1982; Plate, 1995, 2003). In the simplest matrix model, an association of two item-vectors is encoded as the outer product of the two vectors, and stored by summing those outer products into a memory matrix. Alternatively, in a convolution-based model, an association of two item vectors is encoded by applying the convolution operation to the two vectors representing a pair of items, which, itself, results in a vector. In the case of circular convolution (defined below, in Equation 6), the association even has the same dimensionality as the item vectors (Plate, 1995). Those convolutions are then summed into a cumulative memory vector.

Each model mechanism has both strengths and weaknesses (for discussions, see, e.g., Pike, 1984; Plate, 1995). Although we will not definitively decide between these model mechanisms here, we present one line of reasoning that suggests convolution may be neurally plausible. This addresses one particular characteristic of convolution models that has been flagged as a potential weakness. That is, convolution (unlike matrix outer-product) will only work if item representations are "noise-like." This term means that element values are not statistically related to one another; in technical terms, the auto-correlation of values across vector indices must be nearly zero (except for a single value of 1 at lag = 0; this is known as a Kronecker δ vector). Vectors with patterns of values that have this property approximate what is called "white noise." Noise-like representations are typically generated in models by randomly assigning each element of the vector a value drawn from a normal distribution (Plate, 1995), the key point being that each vector element is drawn completely independently from all other element values.

However, if information people remember derives from the natural world, there is no a priori reason to assume that representations of that information will be anything close to noise-like. Consider that naturalistic stimuli (like photographs of the real world) are not noise-like, but in fact, highly auto-correlated. Spe-

This article was published Online First February 13, 2017.

Kenichi Kato and Jeremy B. Caplan, Department of Psychology, University of Alberta.

Supported by the Natural Sciences and Engineering Research Council of Canada.

Correspondence concerning this article should be addressed to Kenichi Kato, Department of Psychology, Biological Sciences Building, University of Alberta, Edmonton, Alberta T6G 2E9, Canada. E-mail: kkato1@ualberta.ca

cifically, naturalistic signals tend to have power spectra of the form, $P(f) = f^{-\alpha}$, also known as "coloured noise" where P and f refer to power (amplitude squared) and frequency, respectively, and $0 < \alpha < 2$ (Field, 1987). White noise would have $\alpha = 0$; in contrast, naturalistic stimuli tend to have lower-frequency components that are much larger (overrepresented) than higher frequencies. Thus, one could criticise convolution as implausible and impractical because it is unsuited to the statistical properties of real-world information. Alternatively, one could ask how the brain generates noise-like representations from information (e.g., stimuli) that contain auto-correlations.

Plate (1995, 2003) suggested a way around this limitation, which Kelly (2010) successfully demonstrated (see also, Kelly, Blostein, & Mewhort, 2013). They started with naturalistic (auto-correlated) stimuli, and then applied a randomly selected permutation to the order of vector dimensions before encoding. This ensured that the vectors would usually be noise-like. If one then applies the inverse permutation after retrieval, this can recover the original (naturalistic) stimulus with very little distortion.

As we show in a later section, perfectly white vectors (where the auto-correlation is fixed to be precisely an identity vector under the convolution operation) are optimally compatible with convolution; indeed, unlike noise-like vectors, a single pair of white-spectrum item vectors is stored and retrieved with zero information-loss. However, it is unlikely that the nervous system uses *precisely* white representations, and it is unclear to us how information could be preserved within such representations.

Drawing inspiration from neuroscience, we propose that the brain's representations of information may already be a kind of compromise between optimality with respect to convolution and preserving the most important features of stimuli. Noise-like representations are often referred to as "decorrelated" representations, because one would otherwise expect representations to inherit the auto-correlated property from real-world signals. They are prevalent throughout the brain. This is most obvious in the early visual processing pathway; retinal ganglion cells have "centre-surround" properties (Srinivasan, Laughlin, & Dubs, 1982). These properties are thought to be because of lateral inhibition (with a broader spread than the more local excitatory connectivity), and has been successfully modelled by the Difference-of-Gaussians function (DoG; Marr & Hildreth, 1980). A DoG is, quite literally, calculated by subtracting a wider Gaussian function (representing the longer-range inhibition) from a narrower Gaussian function (representing the short-range excitation). If a DoG-transform is applied to a visual stimulus, the effect is to enhance edges and reduce intensity levels between edges, producing a kind of outline or cartoon-like visual impression (Figure 9a and 9b). The power spectrum of a DoG function has small power values at low frequencies; thus, acting like a filter, when applied to auto-correlated naturalistic stimuli, a DoG transform also has the effect of somewhat flattening the power spectrum. This counteracts the large power values that are characteristic of naturalistic, auto-correlated stimuli, so the resulting representations are decorrelated (Balboa & Grzywacz, 2000; Field, 1987), and should be closer to optimal for convolution-based memory models. Therefore, we speculated that DoG-transformed natural images might be relatively compatible with convolution, and thus, suffer less distortion than auto-correlated naturalistic images when stored and retrieved in such a model, which may suggest that convolution-based memory is suitable for the kind of representations of information the brain appears to provide.

Decorrelation seems to be a general computational principle in the brain. For example, the lateral geniculate nucleus of the thalamus decorrelates visual information in the temporal domain (Dong & Atick, 1995). This enhances sensitivity to transients, removing the remaining (relatively unchanging) redundant time-course information, but formally, this is quite close to producing noise-like temporal representations of time-varying visual stimuli. Other in vivo recordings in neocortex have found that spike trains exhibit near-zero correlations (Renart et al., 2010), suggesting that decorrelation of representations of information might be quite common throughout the brain. Therefore, the noise-like condition may in fact be satisfied in many pathways in the brain, perhaps via a similar computational mechanism (lateral inhibition), at many different levels of representation. Note that convolution models are typically applied to memory tasks that involve "items" at a high level of abstraction, far from retinal representations. However, retinal-like representations are straightforward to visualize. For this reason, we test the compatibility of convolution-based memory models with neurally realistic, decorrelated representations, using DoG-transformed photographs as a test case, to enable us to evaluate the results not only objectively, but also subjectively. However, because lateral inhibition appears in numerous regions of the brain, we are suggesting that the same principle may apply to the higher cognitive representations that might be present, for example, in medial temporal lobe.

We first explain the problem that auto-correlated naturalistic stimuli pose for convolution models. Then, we show the strengths and weaknesses of noise-like representations, which are commonly used in convolution modelling work, and white representations, which are mathematically optimal for convolution. Finally, we demonstrate how DoG-transformed photographs may represent a middle-ground kind of representation of stimuli wherein important features (edges) can be stored and retrieved with only a small amount of distortion.

The Basic Operation of a Convolution-Based Association-Memory Model

In a convolution-based memory model, an item representation is denoted as an n-dimensional vector:

$$\mathbf{x} = (x_0, x_1, \cdots, x_{n-1}), \tag{1}$$

where a vector is written as a letter or letters with boldface and x_j denotes the value of the j-th element of the vector. The vector can be transformed into the frequency domain by DFT (Discrete Fourier Transform), denoted by FT() and computed as follows:

$$\mathbf{X} = \mathrm{FT}(\mathbf{x}) = (s_0 e^{i\theta_0}, s_1 e^{i\theta_1}, \cdots, s_{n-1} e^{i\theta_{n-1}}), \tag{2}$$

where s_j is the amplitude of the j-th element of the vector in the frequency domain and θ_j is the phase of the j-th element. We follow the convention of using an uppercase letter to denote a vector in the frequency domain corresponding to the (lowercase-letter) vector in the original domain; in our examples, the original domain is spatial. Each element in the frequency domain will be calculated as follows:

$$X_j = \sum_{k=0}^{n-1} x_k e^{-2\pi i j k/n}, \tag{3}$$

where $i = \sqrt{-1}$ and X_j is the j-th element of the vector X. then,

$$s_j = |X_j| \quad \text{and} \quad \theta_j = \mathtt{arg}(X_j), \tag{4}$$

where $|x|$ denotes the absolute value of $x \equiv \sqrt{\text{real}(x)^2 + \text{imag}(x)^2}$, and arg denotes the angle of $x \equiv \tan^{-1}[\text{imag}(x)/\text{real}(x)]$. Memory for an association is produced by the convolution of two item vectors and stored as a new vector with the same dimensionality as the item vectors. In the spatial domain, this is written:

$$\mathbf{z} = \mathbf{x} \circledast \mathbf{y}, \tag{5}$$

where $\mathbf{x}$ and $\mathbf{y}$ are item vectors, $\mathbf{z}$ is the association, and $\circledast$ denotes (circular) convolution. Circular convolution is calculated as follows (Plate, 1995):

$$z_j = \sum_{k=0}^{n-1} x_k y_{(j-k) \bmod n}, \tag{6}$$

where mod denotes modulo. A schematic depiction of the circular convolution calculation through a Latin square is shown in Figure 1a and 1b, as introduced by Kelly (2010) and Kelly et al. (2013).

In the frequency domain, convolution is represented as follows:

$$\begin{aligned} \mathbf{Z} &= \text{FT}(\mathbf{z}) = \text{FT}(\mathbf{x} \circledast \mathbf{y}) = \text{FT}(\mathbf{x}) \odot \text{FT}(\mathbf{y}) = \mathbf{X} \odot \mathbf{Y} \\ &= (s_0 t_0 e^{i(\theta_0+\varphi_0)}, s_1 t_1 e^{i(\theta_1+\varphi_1)}, \cdots, s_{n-1} t_{n-1} e^{i(\theta_{n-1}+\varphi_{n-1})}), \end{aligned} \tag{7}$$

where $\odot$ denotes element-wise multiplication, and

$$\mathbf{Y} = (t_0 e^{i\varphi_0}, t_1 e^{i\varphi_1}, \cdots, t_{n-1} e^{i\varphi_{n-1}}). \tag{8}$$

Correlation, the approximate inverse of convolution, is used to model retrieval of associations via cued-recall, where one item of a pair is given as a memory cue and the model (or participant) has to retrieve its associate. In the spatial domain, correlation, denoted by the operator $\circledcirc\!\!\!\!\#$, is defined as:

$$\mathbf{z} = \mathbf{x} \circledcirc\!\!\!\!\# \mathbf{y} = \text{inv}(\mathbf{x}) \circledast \mathbf{y}, \tag{9}$$

where inv() denotes involution, which reflects the vector elements around the middle-element (Plate, 1995):

$$\text{inv}(\mathbf{x})_j = x_{-j \bmod n}. \tag{10}$$

The Fourier transform of the involution of a vector is the complex conjugate[1] (denoted with superscript *) of the Fourier-transformed original vector:

$$\text{FT}(\text{inv}(\mathbf{x})) = \mathbf{X}^* = (s_0 e^{-i\theta_0}, s_1 e^{-i\theta_1}, \cdots, s_{n-1} e^{-i\theta_{n-1}}). \tag{11}$$

In the frequency domain, correlation is straightforward:

$$\mathbf{Z} = \mathbf{X}^* \odot \mathbf{Y}. \tag{12}$$

Finally, the frequency-domain vector can be converted back to the spatial domain by inverse Fourier transform, denoted FT^{-1}:

$$\mathbf{x} = \text{FT}^{-1}(\mathbf{X}). \tag{13}$$

As in the case of the (discrete) Fourier transform, each element in the spatial domain will be calculated back from the vector in the frequency domain in the inverse Fourier transform as follows:

$$x_j = \frac{1}{n} \sum_{k=0}^{n-1} X_k e^{2\pi i j k/n}. \tag{14}$$

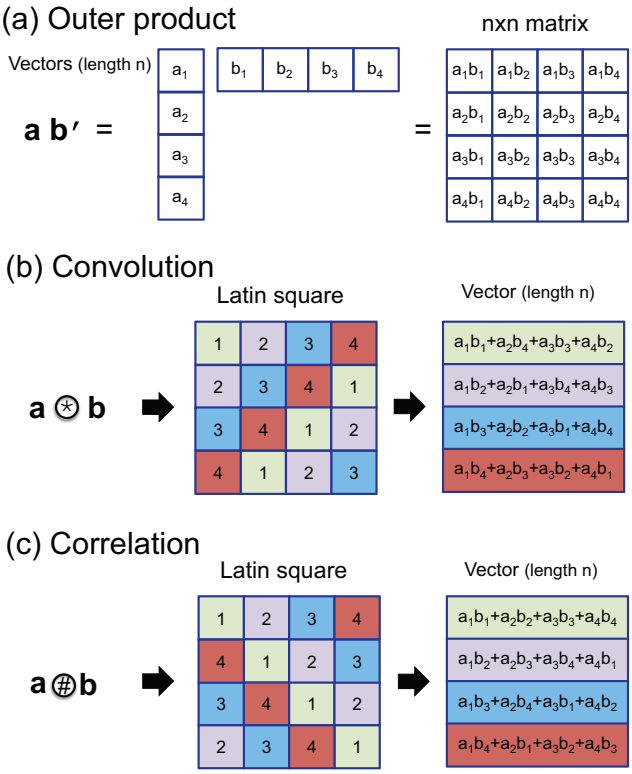


Figure 1. A schematic depiction of the convolution and correlation calculations in the case of two item vectors, **a** and **b**, with four elements each. First, (a) the outer product between the two vectors is computed. Next, (b) for convolution, cells of the outer-product matrix are summed (right) in a pattern that can be visualized as a Latin square (b; middle). (c) Circular correlation can be written by substituting a Latin square that sums over the main diagonals (also wrapping around the matrix). Circular convolution and correlation can be more compactly expressed in the frequency domain (see main text). See the online article for the color version of this figure.

Naturalistic Stimuli as Item Representations in a Convolution Model

The first thought one might have is to simply use the convolution model to store and retrieve real, natural stimuli (Kelly, Blostein, & Mewhort, 2013). Here, stimulus values are used directly as item representations (e.g., real sounds or images). We tested our simple convolution model with a set of 16 photographs taken by the first author, KK, within the city of Edmonton. The resolution was 768 × 768 pixels, originally with 256 gray levels

[1] To avoid possible confusion, note that modellers frequently use * also to denote the convolution operation. Because we use $\circledast$ to denote (circular) convolution, we reserve the uncircled * to denote complex conjugate.

($M = 127$, $SD = 55$). The stimulus set is available to be shared if requested.

The ubiquitous auto-correlated characteristic of natural stimuli causes problems, as follows. Figure 2 displays two photographs: LAMPPOST (that we denote by the vector, **lamppost**), and BRIDGE (**bridge**). To allow us to visualize the results, all item representations are represented as two-dimensional arrays and the convolution and the correlation are calculated by means of a two-dimensional Fourier transform function. In Fourier analysis, a 2D Fourier transform (and its inverse) can be computed in Cartesian coordinates for x and y independently. The math is a trivial extension from 1D to 2D:

$$X_{jk} = \sum_{l=0}^{n-1} \sum_{m=0}^{n-1} x_{lm} e^{-2\pi i(jl+km)/n}, \quad (15)$$

$$x_{jk} = \frac{1}{n^2} \sum_{l=0}^{n-1} \sum_{m=0}^{n-1} X_{lm} e^{2\pi i(jl+km)/n}. \quad (16)$$

For the case of a square matrix, we used Matlab's fft2.m function to implement it. We chose to keep everything in 2D rather than unwrap into 1D, then rewrap into 2D (as Kelly has done), because we felt this way, the statistical properties of the photographs would be preserved. The unwrapping might introduce artificial "edges"—sharp transitions at the boundaries of the rows of the image and make things needlessly complicated. Because the mathematics of 2D Fourier transform is so close to the mathematics of 1D Fourier transform, we remained in 2D for the demonstrations. The association (convolution) of the two images is displayed in the top-right, and the retrieved LAMPPOST image (**lamppost**$_r$) from the association, by applying **bridge** as a cue, is in the bottom-right. The retrieved image looks very blurry, which is precisely because of the auto-correlation: Auto-correlation indicates more power at lower frequencies. These low frequencies thus, dominate the convolution and correlation operations and have an effect like smoothing the image. Typically, convolution models are assessed by computing the similarity (normalized dot product) between the retrieved vector and possible response vectors; the higher the dot product, the more likely an item is to be produced as a response. Thus, to quantitatively assess the performance of the model with these kinds of stimuli, we calculate the normalized dot product between the retrieved image and the correct (target) and incorrect (other candidate) images. The similarity (normalized dot product, equivalent to the cosine of the angle between a pair of vectors) between the original and retrieved LAMPPOST images, 0.459, is larger than the value between the original BRIDGE and the retrieved LAMPPOST images, 0.236. Thus, quantitatively, the retrieved LAMPPOST image can be effectively distinguished as **lamppost** from **bridge**, but information in the middle- to high-frequency ranges, that may be just as diagnostic of stimuli, is degraded.

Item Representations With Randomly Assigned Values

To avoid the smearing effect because of naturalistic stimuli, most modellers working with convolution use item vectors consisting of values drawn at random. Drawing values at random ensures that in general, there will be very little auto-correlation in item representations. Typically (Murdock, 1982; Plate, 1995), val-

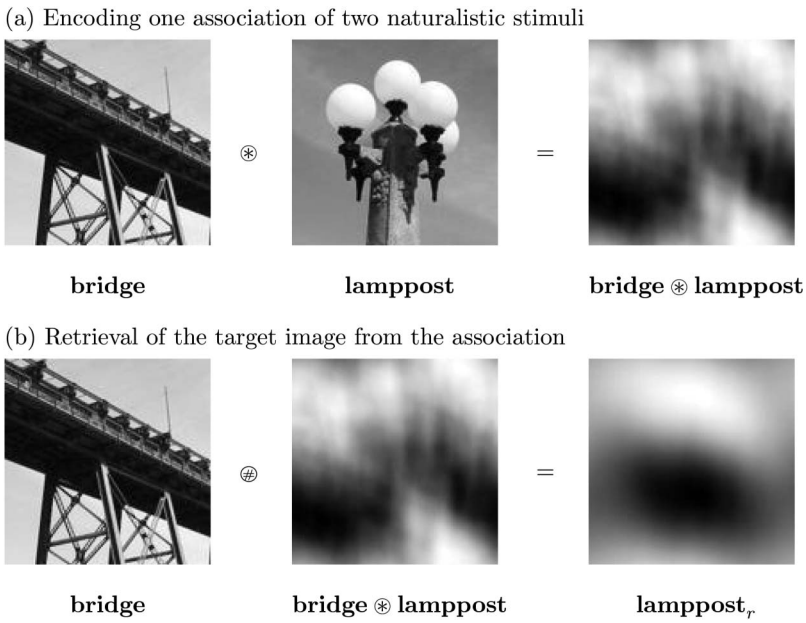


Figure 2. An example of encoding one association of two items with naturalistic stimuli (128 × 128 pixels) and retrieval of the target vector (**lamppost**) from the association. The retrieved image (**lamppost**$_r$) looks quite blurred. The similarity (cosine) between the two is 0.459, which is still distinguishable from the similarity between the cue (**bridge**) and the retrieved vectors, 0.236. To increase visibility, all images presented here are normalized to the length one and contrast-enhanced by applying a sigmoid function to pixel-values, x: $1/(1 + \mathbf{exp}^{-100x})$.

ues are independent, identically distributed (i.i.d.), meaning they are drawn from Gaussian distribution with $\mu = 0$ and $\sigma = 1/\sqrt{n}$ (where n is the number of the elements of the vector), and each value is drawn at random, without any systematic relationship to other values. The mean and variance ensure that vectors are approximately normalized (the expectation of the dot product of the vector with itself is equal to 1; $E[\mathbf{x}\cdot\mathbf{x}] = 1$).

Figure 3 shows an example of storing one association, **mouse** ⊛**cat**. When cued with **mouse**, the model retrieves a vector which we denote $\mathbf{cat}_r$. It is hard to see, visually, the similarity of the retrieved pattern, $\mathbf{cat}_r$, to the original target pattern, **cat**, because the values are arbitrary. However, when measured quantitatively, with the normalized dot product, the similarities between the original and retrieved vectors are sufficiently distinguishable. In this case, with 8 × 8 pixel "vectors" (see Figure 3), the similarity between the original and retrieved item (**cat** and $\mathbf{cat}_r$) is 0.649. This is considerably larger than, for example, the value between the original MOUSE, **mouse**, and the retrieved CAT, $\mathbf{cat}_r$, −0.034 (**mouse** and **cat** were composed of random values chosen completely independently of one another, so this will tend to be true for other items in the "lexicon" as well). For larger dimensionality, in the case of 768 × 768 pixel items (see Table 1), the values are 0.707 ($1/\sqrt{2}$) and close to zero, respectively.

This shows why item-vectors drawn from random values can function effectively in a convolution model. However, one still has to explain how such representations might have been derived from the original stimulus properties. As Figure 3 shows, even though the similarity between the retrieved cat, $\mathbf{cat}_r$, and the original **cat** has sufficiently high value, 0.649, it is very difficult to subjectively see these two images are similar. To appreciate this, we plot the vector-element values of the original vector against those for the retrieved vector. In Figure 4a, plotting **cat** against $\mathbf{cat}_r$, one can see the relationship between the two sets of vector-element values, whereas in Figure 4b, plotting **mouse** against $\mathbf{cat}_r$, there is no such relationship.

Representations "Protected" by Random Permutations

Kelly et al.'s (2013) approach was to apply a random permutation to the vector dimensions of items before encoding them. The random permutation effectively protects a naturalistic stimulus from smearing because of its own auto-correlation because a random shuffle of stimulus values is very likely to approximate a white (uncorrelated) vector. In their method, two permuted item representations are convolved, and then the retrieved vector is permuted back by applying the inverse permutation. Figure 5 and Table 1 show an example of the random-permutation method. In the simulation, each item was assigned a different permutation table. The random-permutation approach thus, addresses the problem of auto-correlated stimulus dimensions by ensuring that the convolution and correlation operations act on vectors that have minimal auto-correlation. Thus, a convolution acting on vectors that are protected by random permutations can perform as effectively as convolution with noise-like representations, but while preserving all information present in the original stimuli. This approach may have important applied uses. However, the limita-

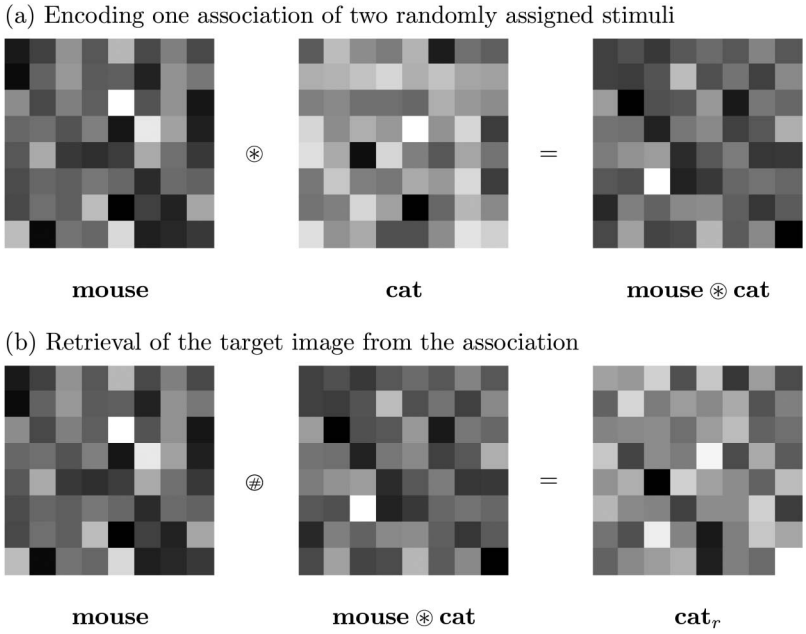


Figure 3. An example of encoding one association of two items with randomly assigned values (8 × 8 pixels) and retrieval of the target vector (**cat**) from the association. The original target (**cat**) image looks different from the retrieved target ($\mathbf{cat}_r$). However, the similarity (cosine) between the two is 0.649, which is sufficiently distinguishable from the similarity between the cue (**mouse**) and the retrieved vectors, −0.034. In the case of 768 × 768 pixel images, the similarity (cosine) between the target vector and the retrieved vector is 0.707, which is very close to theoretical value for Gaussian random values ($1/\sqrt{2} = 0.707$) and easily distinguishable from the similarity between the cue vector and the retrieved vector, 0.002.

Table 1
Average Similarities Between Retrieved and the Candidate Vectors Over Five Representation Types

N	Item type	Mean/Max	Natural	DoG	Permute	Random	White
1	target	mean	0.55	0.47	0.71	0.71	1.0
	other candidates	mean	0.013	−0.00041	0.00014	0.00011	−0.00021
		max	0.43	0.0026	0.0046	0.055	0.0028
2	target	mean	0.42	0.40	0.58	0.58	0.71
	other candidates	mean	0.0092	−0.0010	0.00015	0.00011	−0.00020
		max	0.42	0.0026	0.0040	0.051	0.0028
4	target	mean	0.30	0.32	0.45	0.45	0.50
	other candidates	mean	0.0061	−0.0013	0.00014	0.000074	−0.00018
		max	0.42	0.0026	0.0035	0.051	0.0027
8	target	mean	0.21	0.25	0.33	0.33	0.35
	other candidates	mean	0.0076	−0.0013	0.00014	0.00011	−0.00015
		max	0.43	0.0026	0.0031	0.048	0.0027

Note. Target items were excluded from candidates. The parameters for DoG-transform were $\sigma_1 = 1$, $\sigma_2 = 4$. Natural = naturalistic images (the original photographs); DoG = natural images after a Differences-of-Gaussian transform; permute = natural images "protected" by permutation; random = "noise-like" images produced by randomly selecting numbers from i.i.d., distributions; white = natural images after precise whitening. The candidate-mean is the mean of the similarities between retrieved vector and all candidates except the targets over all possible combinations in 16 items whereas the candidate-max is the mean of the maximum similarities between retrieved vector and 15 candidates (except the target) for each association. Even though the mean similarities between the target and the retrieved images in natural images are greater than those in DoG-transformed images in the case of one and two associations, the maximum similarities for other candidates in natural images are very large so the probability of correct choice will be greater In DoG-transformed representation than in natural images.

tion of this approach for developing a realistic model of human memory, as acknowledged by Kelly et al. (2013), is that one needs to explain how the random permutation is produced, and has to preserve the precise permutation selected to decode the item after the correlation step.

Whitened Naturalistic Stimuli Would Be Optimal

There is, in fact, a known, optimal type of vector representation. Based on the convolution and correlation operations, retrieval of the target vector from one association vector is represented as follows. In the spatial domain, retrieval from a memory trace, correlating with the cue-item vector, is

$$\mathbf{x_r} = \mathbf{y} \circledast (\mathbf{x} \circledast \mathbf{y}) = (\mathbf{y} \circledast \mathbf{y}) \circledast \mathbf{x}, \tag{17}$$

where x_r denotes the retrieved vector. In the frequency domain, the retrieval of the target vector $\mathbf{X}_r$ is defined:

$$\begin{aligned}\mathbf{X_r} &= \mathbf{Y}^* \odot \mathbf{X} \odot \mathbf{Y} \\ &= (s_0\, t_0^2\, e^{i\theta_0},\ s_1\, t_1^2\, e^{i\theta_1},\ \cdots,\ s_{n-1}\, t_{n-1}^2\, e^{i\theta_{n-1}}).\end{aligned} \tag{18}$$

The amplitude, t^2_j, of the j-th element of the retrieved vector originates from the cue vector and this produces a distortion in the retrieved vector. Figure 6 illustrates the distortion in the frequency domain. Figure 6a shows the power spectrum of a vector con-

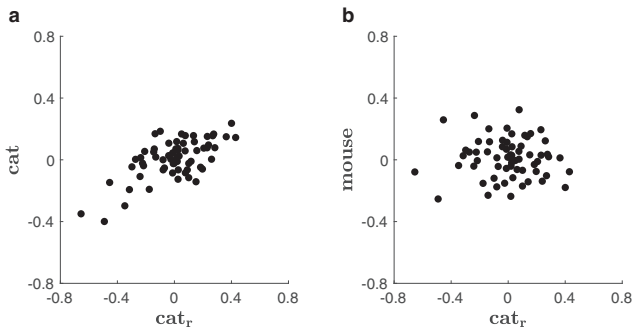


Figure 4. Because noise-like vectors are hard to evaluate visually (Figure 3, we plot the vector element values for one item against another (each dot in the scatter plot), for (a) original cat vector compared with the retrieved vector; (b) original mouse vector compared with the retrieved vector.

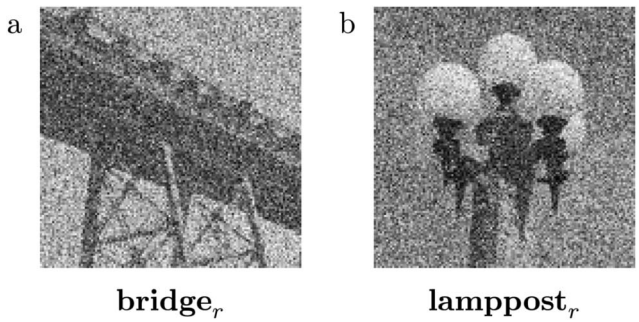


Figure 5. Examples of naturalistic stimuli (BRIDGE and LAMPPOST images, 128 × 128 pixels) after being randomly permuted, stored, and then retrieved with cued recall and had the permutation corrected. Because the permutation is selected at random, the retrieved images appear as though they had noise added to them. However, they are still quite recognizable. The similarity (cosine) between the target vector (**lamppost**) and the retrieved vector is 0.708, which is very close to theoretical value for Gaussian random values ($1/\sqrt{2} = 0.707$) and easily distinguishable from the similarity between the cue vector (**bridge**) and the retrieved vector, 0.002. Visibility was enhanced as in Figure 2.

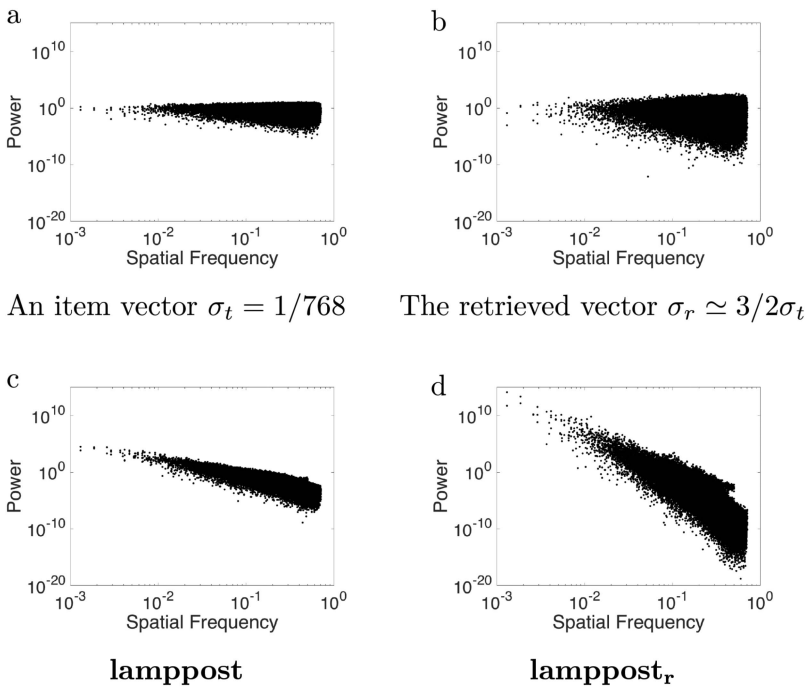


Figure 6. Examples of power spectra of a randomly assigned vector and naturalistic stimuli (both 768 × 768 pixels). The average of the power of the randomly assigned vector is about 1, whereas the power spectrum of the naturalistic stimuli (**lamppost**) is inversely proportional to the squared frequency. The power spectra of retrieved items have larger σ for both representation.

structed from random (i.i.d.) values, and Figure 6b shows the power spectrum after being stored and then retrieved from a convolution model. In the random-value representation, the power spectrum is approximately flat across frequencies, as expected. Even though the amplitudes of cue vectors increase the deviation in the retrieved vector, the nature of the flat distribution is not affected by the unwanted term, t^2_j. Figure 6c and 6d show the same for one naturalistic stimulus (LAMPPOST). In the naturalistic image, the power spectrum starts out sloped, and this slope becomes even larger after encoding (via convolution) and retrieval (correlation); $1/f^{\alpha}$ in original becomes roughly $1/f^{3\alpha}$ in the retrieved vector because the amplitude of the retrieved vector is $S_{r.j} = S_j t^2_j$ in the frequency domain.

To achieve optimal retrieval, all the amplitude values of the cue vector must be set to 1, which is equivalent to "whitening" the vector (all frequencies have the same amplitude):

$$t_0 = t_1 = \cdots = t_{n-1} = 1. \tag{19}$$

The whitened item vector in the frequency domain can be written:

$$\text{white}(\mathbf{Y}) = (e^{i\varphi_0}, e^{i\varphi_1}, \cdots, e^{i\varphi_{n-1}}). \tag{20}$$

The same logic can be applied to retrieval of the other vector, **Y**, so the amplitude values of white(**X**) must be 1 as well (Caplan, 2011):

$$\text{white}(\mathbf{X}) = (e^{i\theta_0}, e^{i\theta_1}, \cdots, e^{i\theta_{n-1}}). \tag{21}$$

In the case of whitened vectors, an association can be written in the frequency domain as:

$$\begin{aligned}\text{white}(\mathbf{Z}) &= \text{white}(\mathbf{X}) \odot \text{white}(\mathbf{Y}) \\ &= (e^{i(\theta_0+\varphi_0)}, e^{i(\theta_1+\varphi_1)}, \cdots, e^{i(\theta_{n-1}+\varphi_{n-1})}).\end{aligned} \tag{22}$$

Then correlation produces perfect retrieval, in the case of one stored association:

$$\begin{aligned}\text{white}(\mathbf{X})_r &= \text{white}(\mathbf{Y})^* \odot \text{white}(\mathbf{Z}) \\ &= \text{white}(\mathbf{Y})^* \odot \text{white}(\mathbf{X}) \odot \text{white}(\mathbf{Y}) = \text{white}(\mathbf{X})\end{aligned} \tag{23}$$

Figure 7 shows an example where the naturalistic stimuli are whitened before encoding an association, and before using an item as a retrieval cue. The retrieved image from an association of two whitened naturalistic stimuli is exactly equal to the target (LAMPPOST) image (see Table 1). For the case of a single association, whitened representations lead to perfect retrieval in cued recall. However, the flat distribution in the power spectrum makes the image appear noisy because of the relatively greater presence of high-frequency components, which has the subjective appearance of noise.

Difference of Gaussians

The DoG filter was initially introduced as a model of lateral inhibition in the early visual system (centre-surround cells in the retinal ganglion), and thought to play an important role in edge detection (Marr & Hildreth, 1980). A DoG is simply the difference of two Gaussian functions with different variances (see Figure 8). Convolving a DoG function with a natural image enhances the medium frequency range and emphasizes edges.

With a convolution model, although a flat power spectrum is optimal, even when imperfect, it is also the case that the flatter the

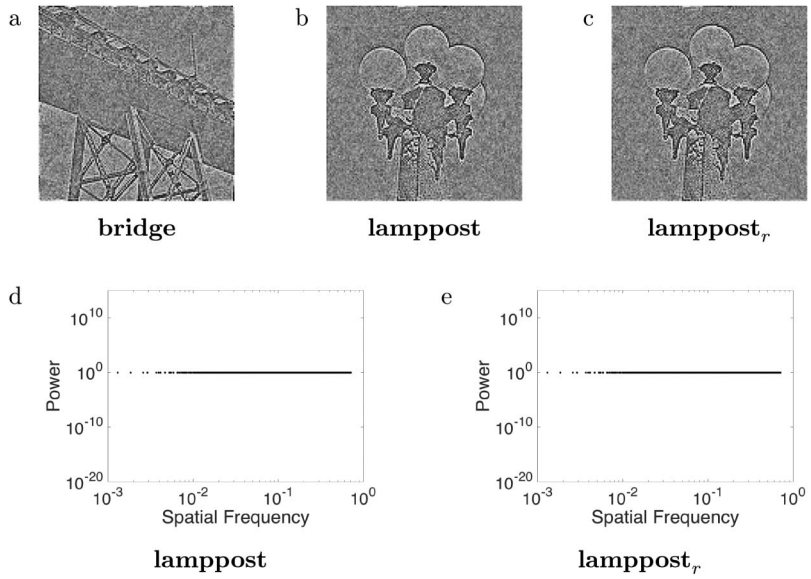


Figure 7. Examples of whitened naturalistic stimuli (BRIDGE and LAMPPOST images, 128 × 128 pixels). The retrieved LAMPPOST image is exactly equal to the target LAMPPOST image. Perfect retrieval is achieved by the flat and no deviation in the power spectrum. We present the images downsampled to 128 × 128 pixels because the original, higher-resolution (768 × 768 pixels) images were harder to evaluate because of distortion when embedded within the document. The similarity (cosine) between the target vector (**lamppost**) and the retrieved vector is 1.000, which corresponds that the retrieved vector is identical to the target vector and easily distinguishable from the similarity between the cue vector (**bridge**) and the retrieved vector, 0.000301. Visibility was enhanced as in Figure 2.

power spectrum of item vector is, the less distortion in the retrieved vector from the association will be. The power spectrum of a DoG-filter shows how DoG-transform works with naturalistic stimuli with 1/*f* distribution (Figure 9e). Because convolution is element-wise multiplication in the frequency domain, the increasing amplitude with frequency counteracts the high power values at low frequencies, partly flattening the 1/*f* form of the spectrum. Figure 9 and Table 1 show examples of the DoG filter applied to naturalistic stimuli. The DoG-transformed ($\sigma_1 = 0.3$, $\sigma_2 = 1.0$, for 128 × 128 images) images have emphasised edges. The retrieved images are noisy, but are improved compared with naturalistic images retrieved from a convolution model (see Figure 2), because the power spectrum is now almost flat in the lower frequency range. The similarity between the original target LAMPPOST and retrieved LAMPPOST images, 0.463, is much larger than that of the cue (**bridge**), −0.027, which suggests the target would be sufficiently distinguishable from other candidate items. When the DoG-filtered images are compared with the original naturalistic images, one can see that, rather than DoG-transform improving the similarity between a target and the retrieved item, it *reduces* the similarity between the retrieved target and the other (unrelated) item vectors, including the cue itself (see entries for $N = 1$ in Table 1).

Memory for Multiple Associations

Thus far, we have examined cases involving memory of only a single association. Naturally, to be useful, a memory system must

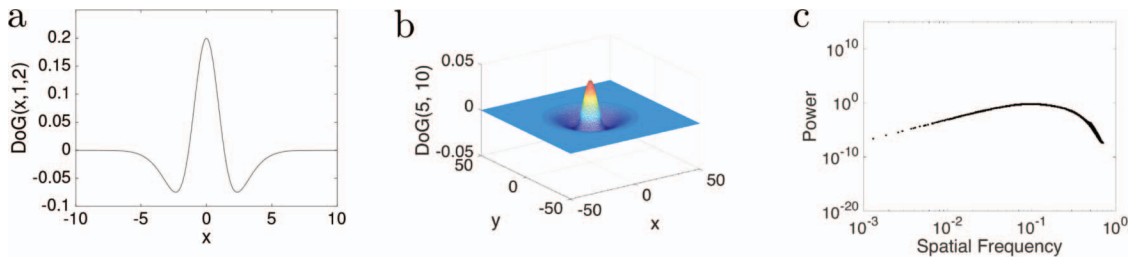


Figure 8. Difference-of-Gaussians functions (DoG). (a) A 1D DoG function ($\sigma_1 = 1$, $\sigma_2 = 2$), (b) a 2D DoG function ($\sigma_1 = 6$, $\sigma_2 = 10$). DoG: DoG is computed from two Gaussian functions with $M = 0$. (c) A power spectrum of DoG filter ($\sigma_1 = 1$, $\sigma_2 = 4$), equivalent to ($\sigma_1 = 0.3$, $\sigma_2 = 1.0$) with the 128 × 128 demonstration images. See the online article for the color version of this figure.

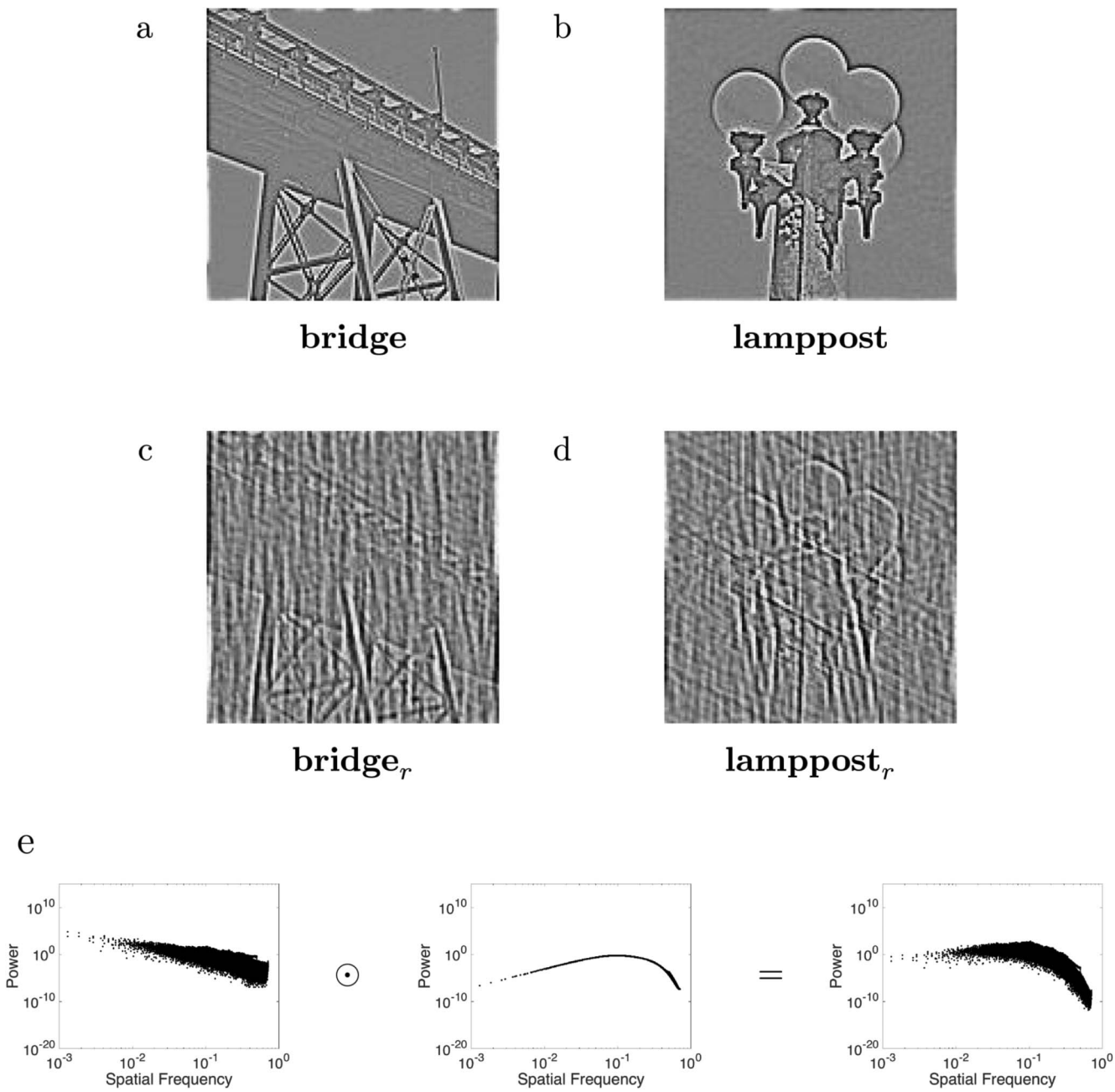


Figure 9. Applying Difference-of-Gaussians (DoG) to naturalistic images as item representations. (a, b) DoG-transformed BRIDGE and LAMPPOST images (128 × 128 pixels, $\sigma_1 = 0.3$, $\sigma_2 = 1.0$), (c, d) The retrieved images from the association of the two DoG-transformed images. The similarity (cosine) between the target vector (**lamppost**) and the retrieved vector is 0.463, which moderately distinguishable from the similarity between the cue vector (**bridge**) and the retrieved vector, −0.027. Visibility was enhanced as in Figure 2. (e) The power spectrum of DoG-transformed BRIDGE image (768 × 768 pixels).

be able to store and retrieve more than a single association. In convolution models, multiple associations are usually stored in a single memory trace by superposition— element-wise addition. As in the case of single association, a target vector is retrieved by correlating the corresponding cue vector to the summed memory vector. For example, for a model with two associations stored,

$$\mathbf{a}_r = \mathbf{b} \circledast (\mathbf{a} \circledast \mathbf{b} + \mathbf{c} \circledast \mathbf{d}) = (\mathbf{b} \circledast \mathbf{b}) \circledast \mathbf{a} + \mathbf{b} \circledast \mathbf{c} \circledast \mathbf{d}, \tag{24}$$

where $\mathbf{a}_r$ denotes the retrieved vector. Note that in a full-fledged model such as TODAM (Murdock, 1982), each association term would be multiplied by a scalar representing an encoding strength,

which could be free to vary across associations; for simplicity, we assume all pairs are stored with equal strength. Because multiple associations are stored additively, the similarity (normalized dot product, or cosine between $\mathbf{a}_r$ and $\mathbf{a}$) must decrease with number of stored associations (Figure 10 and Table 1). The retrieved images are increasingly degraded as the number of associations, N, increases, and the similarity values, correspondingly, decrease with increasing N. Nevertheless, the retrieved image is still recognizable and the similarity values are larger than those of other candidates, which is what ultimately determines memory success for these models.

Figure 11 shows how superposition degrades the retrieved vector when the number of associations, N increases in the case of DoG-transformed representations. The greater the N, the more degraded the retrieved image is. For the purpose of comparison, the same set of images with permuted and whitened item representations are shown in Figures 12 and 13, respectively. The retrieved images over all associations in both permuted and whitened representations are less degraded than those of DoG-transformed representations.

The relation between the number of associations, N, and the probability that the target image is correctly selected, is shown in Figure 14. The probability of correct choice was calculated with a choice rule suggested by Luce (1959):

$$P(i) = \frac{S(i)^{\gamma}}{\sum_j S(j)^{\gamma}}, \tag{25}$$

where $P(i)$ is the probability the i-th candidate is selected, $S(i)$ is a strength calculated with similarity between the retrieved and the candidate vectors, the denominator is the sum of the strengths of all possible candidates, and γ is a constant. $S(i)$ values were truncated at zero, so items with negative strength had zero probability of being sampled.

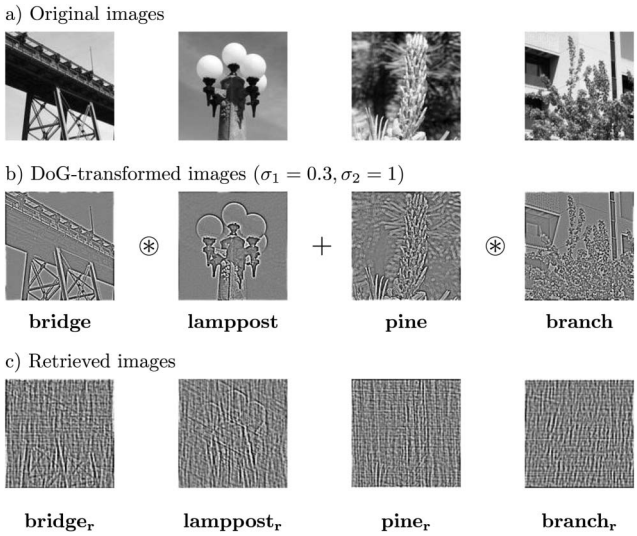


Figure 10. Superposition of two associations of Difference-of-Gaussians (DoG) transformed naturalistic stimuli (128 × 128 pixels). (a) Original images, (b) DoG-transformed images ($\sigma_1 = 0.3$, $\sigma_2 = 1$), (c) The retrieved images from the superposition of the two associations. Visibility was enhanced as in Figure 2.

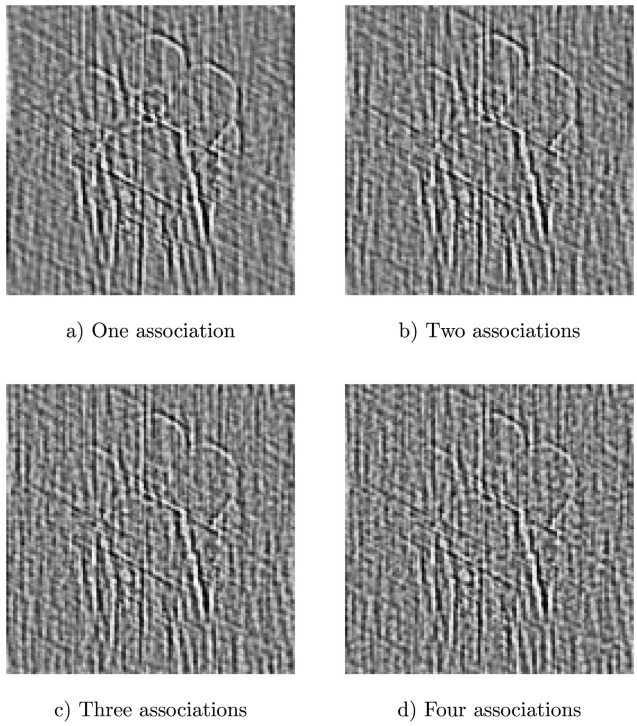
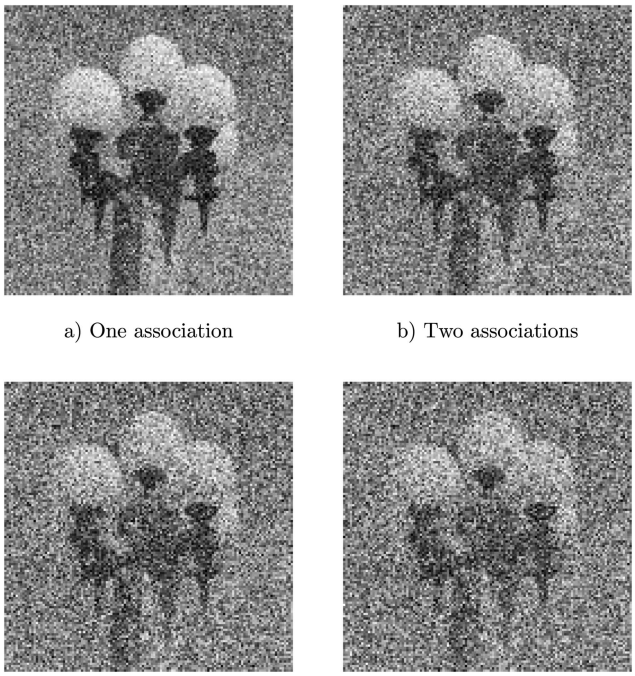


Figure 11. The image of LAMPPOST (128 × 128 pixels) after being retrieved from memories containing several (1–4) associations of Difference-of-Gaussians (DoG) transformed naturalistic stimuli. We present the images downsampled to 128 × 128 pixels because higher-resolution (768 × 768) images were harder to evaluate because of distortion when embedded within the document. The σ_1 and σ_2 in the DoG function were modified to 0.3 and 1, respectively, to optimise the recognisability in the retrieved images. Visibility was enhanced as in Figure 2.

In the simulation, 16 images (original photographs taken by the first author) were used as the item vectors, except for the case of random representations, which were produced with randomly assigned values drawn from a Gaussian distribution with $\mu = 0$ and $var = 1/n$ where n is the number of elements in a vector. All item vectors were normalised to length 1 and mean-centered to 0. Each value of probability correct was the mean of all associations of 16 items, which included all possible combinations of items, excluding the auto-association (association of the item with itself). Superposition was simulated only for specific combinations. For example, if $\mathbf{a} \circledast \mathbf{b} + \mathbf{c} \circledast \mathbf{d}$ was calculated, $\mathbf{a} \circledast \mathbf{b} + \mathbf{e} \circledast \mathbf{f}$ was not calculated to test the association **a** and **b** items but all associations, $\mathbf{a} \circledast \mathbf{b}$, $\mathbf{a} \circledast \mathbf{c}$, $\mathbf{a} \circledast \mathbf{d}$ and so on were calculated and averaged over all combinations. Figure 14a plots probability correct using $\gamma = 1$. As expected, the whitened representations showed the best performance and the naturalistic representations were worst. DoG-transformed representations performed moderately well: clearly not optimal, but substantially better than naturalistic representations. Figure 14b plots the results with $\gamma = 2$, to see what happens if accuracy increases, and shows the same basic pattern as with $\gamma = 1$.

The Luce Choice rule is not the only way to model the final decision process. Therefore, we also plot the maximum similarities for competitor items in Figure 15. As expected, the whitened



a) One association b) Two associations

c) Three associations d) Four associations

Figure 12. The image of LAMPPOST (128 × 128 pixels) after being retrieved from memories containing several (1–4) associations of permuted naturalistic stimuli. The images are scrambled back after the retrieval. The size of the images were reduced from 768 × 768 pixels to 128 × 128 pixels for the purpose of subjective recognisability. Visibility was enhanced as in Figure 2.

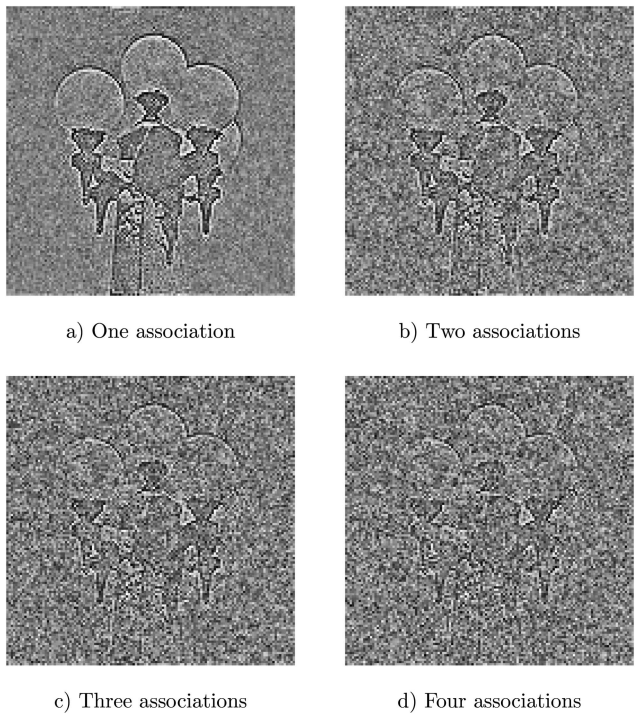
representations showed the highest similarities over all number of associations, and the permuted and the randomly assigned representations showed the second performance. All three representation types showed almost zero similarities between the retrieved vectors and the competitors even in the maxima. In contrast, the DoG-transformed representation did not show better performance than the original naturalistic stimuli in the similarities between the retrieved and the target vectors. However, the maximum similarities of the competitors were very high and comparable with the targets in the original naturalistic stimuli whereas those for the DoG-transformed representations stayed low. Therefore, the DoG-transformed representations can be expected to perform better than the original naturalistic stimuli in the clean-up process.

A few final points should be noted. (a) The similarities of the competitors stayed the same with increasing the number of associations. (b) The advantage of the whitened representations over the permuted and the random representations decreased with the number of associations and disappeared with the eight associations. Therefore, there might be no specific reason to support the whitened representation when the number of associations to be stored is large enough.

Discussion

Convolution-based memory models have successfully explained a very broad range of human-memory empirical phenomena (e.g., Kahana, 2012; Murdock, 1982; Plate, 1995). One of the weaknesses of convolution-based models has been the constraint that item representations had to be noise-like, whereas naturalistic stimuli have long been known to contain large auto-correlations, because of their coloured-noise power spectra. Matrix models, an alternative approach to association-memory, do not share this limitation. Although we cannot resolve the debate about which class of model is more neurally plausible (see, e.g., Murdock, 1985; Pike, 1984; Plate, 1995), we contend that the demonstrations we present here suggest that far from being a limitation, the requirement of decorrelated item representations is (approximately) satisfied by the kinds of representations that are apparently already present in the brain, even as early as the retinal ganglion. The performance of DoG-transformed naturalistic images was, unsurprisingly, not optimal, and in particular, a high-frequency distortion is plainly visible in the images retrieved from the model (Figures 9, 10, and 11). However, the DoG-transformed stimuli were recognizable after having been encoded and retrieved in a convolution model, even with multiple associations stored. Quantitatively, the DoG-transformed stimuli always outperformed the untransformed, naturalistic stimuli. Although we argue that DoG-transformed stimuli are moderately compatible with convolution, it is possible that association-memory in real brains is not precisely convolution, but some modified variant of convolution that could conceivably be better matched to the DoG-transform specifically.

Of the three nonnaturalistic representations we considered, whitened representations are quite obviously (mathematically) optimal. The comparison, in the retrieved images from memories with multiple associations, between DoG-transformed (see Figure 11)



a) One association b) Two associations

c) Three associations d) Four associations

Figure 13. The image of LAMPPOST (128 × 128 pixels) after being retrieved from memories containing several (1–4) associations of whitened naturalistic stimuli. Visibility was enhanced as in Figure 2.

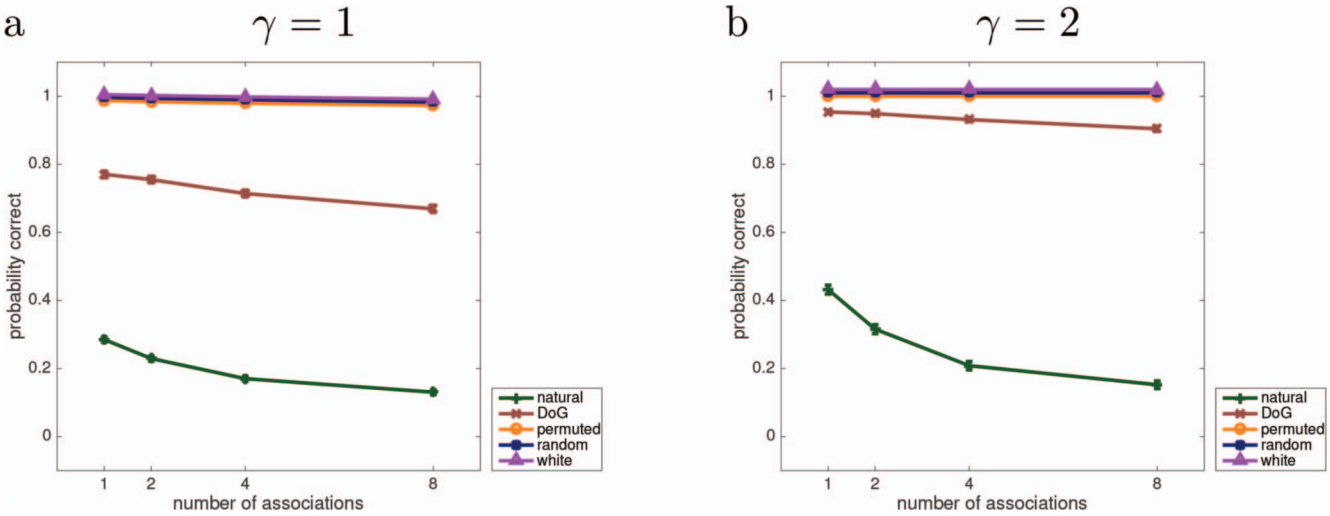


Figure 14. Probability correct as a function of number of stored associations, for each type of representation: natural (raw image), Difference-of-Gaussians (DoG; those natural stimuli after being DoG-transformed), permuted (stored after applying a random permutation and with the inverse permutation applied after retrieval), random (i.i.d. values, the "noise-like" representations typically used in convolution-model research), and white (the naturalistic images after precise whitening), using 16 items and a probabilistic choice rule, with $\gamma = 1$ (a) and $\gamma = 2$ (b); see text. The error bars are the *SEM*, but are small, and in some cases, smaller than the data-point markers. The random-representations curve was slightly shifted upward to avoid it and the permuted representations curve from occluded one another, because these two representation types have almost exactly the same values. See the online article for the color version of this figure.

and whitened naturalistic stimuli (see Figure 13) also shows the advantage for whitened representations visually. However, precise whitening strikes us as too precise to expect of real neuronal networks. Noise-like representations are popular among convolution modellers, but they raise the further question of how one maps stimulus- or item-information onto a set of noise-like vectors. Vectors scrambled by random permutations (Kelly, 2010; Kelly et al., 2013) preserve the original naturalistic features quite well and performed as well as randomly assigned item representations (Figures 14 and 15), but they raise further questions: how is the "random" permutation selected, is it plausible that the random permutation is stored and available when the unscrambling is required? And, as with exact-whitening, we do not know of any evidence that the random-permutation mechanism takes place in the brain.

Rather, we view approximate decorrelation as carried out via lateral inhibition as a general computational feature of the brain, quite likely for other reasons such as energy-efficient coding (Balboa & Grzywacz, 2000; Dong & Atick, 1995; Field, 1987; Renart et al., 2010; Tanaka, 2003), which then happens to be compatible with convolution. Our simulations show that the deviation of DoG-transformed stimuli from noise-like is nonetheless sufficient to produce good performance in a convolution model.

One intriguing possibility is that the entorhinal cortex may be set up to compute convolution (or some close approximation to convolution). As explained earlier, convolution, in the frequency domain, is simple element-wise multiplication. The so-called "grid cells" in medial entorhinal cortex (Fyhn, Molden, Witter, Moser, & Moser, 2004; Hafting, Fyhn, Molden, Moser, & Moser, 2005) have 2-dimensional, periodic place fields that vary across a large range of spatial frequencies. Numerous researchers, starting with Solstad, Moser, and Einevoll (2006), have noted that this resembles a Fourier basis set. What follows from this is that the activity of a medial entorhinal cell could be viewed as a Fourier coefficient (for a given frequency and phase). Then, an association between two such "item" vectors in the frequency domain, could be learned in a convolution-like manner via direct multiplication (Caplan, 2011).

Our use of images was to enable us to visualize the performance of the model and the degradation of the stimuli. However, we are not proposing that human brains store and retrieve DoG-transformed versions of retinal images. Presumably, brain regions, such as hippocampus, entorhinal cortex and other medial-temporal-lobe regions store and retrieve associations in episodic memory. However, it is still unknown what the exact form of the representations of this higher-order information is. What we propose is that, at every stage in the brain, the inputs to a region contain auto-correlations, with approximately coloured-noise form, and that via lateral inhibition, each region removes much of this auto-correlation at that particular level of representation. Thus, if not literally, we propose that conceptually, our results here demonstrate that the medial temporal lobe could, relatively safely, apply convolution to store associations, followed by correlation to retrieve them (or some operations approximating convolution and correlation).

There are, in fact, models of semantic memory, which may be pertinent here. We examined the degree of auto-correlation of item representations across three major semantic representation models, Latent Semantic Analysis (LSA; Landauer & Dumais, 1997), Hyperspace Analogue to Language (HAL; Lund & Burgess, 1996;

Burgess, 1998), and Bound Encoding of the AGgregate Language Environment (BEAGLE; Jones & Mewhort, 2007), with sample data sets. The TASA, EN_100k, and blogs_beagle data sets were used for LSA, HAL, and BEAGLE models, respectively. The TASA dataset consists of 92,393 English words with 300 elements each, EN_100k dataset 100,000 English words with 300 elements each, blogs_beagle dataset 103,599 German words with 1,024 elements each, respectively, as computed and posted by Günther (2015). Figure 16 plots the power spectra of item representations of the three models, comparing to that for 32 128 × 128 natural images (photos), 16 of which we used in the other demonstrations in the manuscript. Fitting a linear regression to the spectra in log/log coordinates, the mean α values (*SD*) were 0.0042 (0.12), −0.0508 (0.19), and −0.0053 (0.06) for LSA, HAL, and BEAGLE, respectively. Thus, all the α values of the semantic models are close to zero, which corresponds to no auto-correlation (approximately white representations), in contrast to α = 3.1 (0.44) in the natural images, reflecting coloured-noise as expected.

However, the ordering of the dimensions in all three methods was never an explicit part of the algorithm, but rather, ranked based on considerations such as proportion of covariance explained. One would still need to look for evidence about how semantic dimensions are laid out topographically in the brain. At least as far as inferotemporal cortex, which appears to respond to (or perhaps even represent) complex visual objects at a relatively high level (invariant to translation, rotation, luminance, etc.), neurons seem to be organized in columns, whereby nearby neurons tend to respond to similar objects (Tanaka, 1993, 2003; Wang, Tanaka, & Tanifuji, 1996). This suggests that the problem of auto-correlated representations exists even at high levels of representation. Tanaka (2003) even suggested the columnar organisation could indicate the presence of local excitation and lateral inhibition. This might support decorrelation similar to what we demonstrated, but at a higher level of representation.

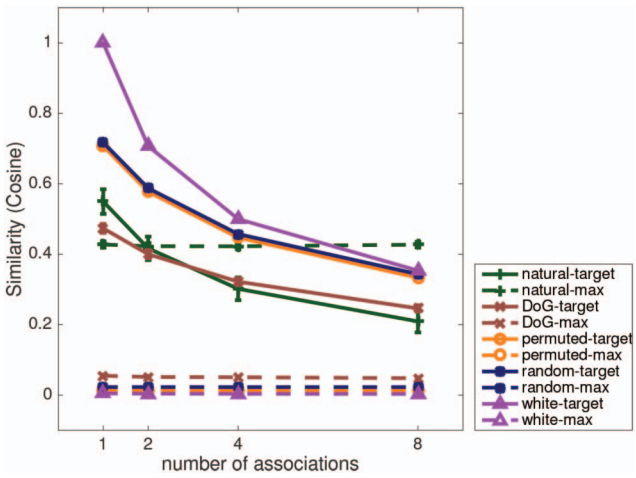


Figure 15. Mean similarities (cosine values) between retrieved and candidate vectors as a function of number of stored associations, for each type of representation: natural (raw image), Difference-of-Gaussians (DoG, those natural stimuli after being DoG-transformed), permuted (stored after applying a random permutation and with the inverse permutation applied after retrieval), random (i.i.d. values, the "noise-like" representations typically used in convolution-model research), and white (the naturalistic images after precise whitening), using 16 items; see text. The error bars are the *SEM*, but are small, and in some cases, smaller than the data-point markers. The solid lines correspond to the similarities between the retrieved and the target vectors whereas the dashed lines correspond to the maximum similarities between the retrieved and the other candidate vectors. For the visibility, the points of the permuted and the randomly assigned representations were artificially shifted upward slightly. See the online article for the color version of this figure.

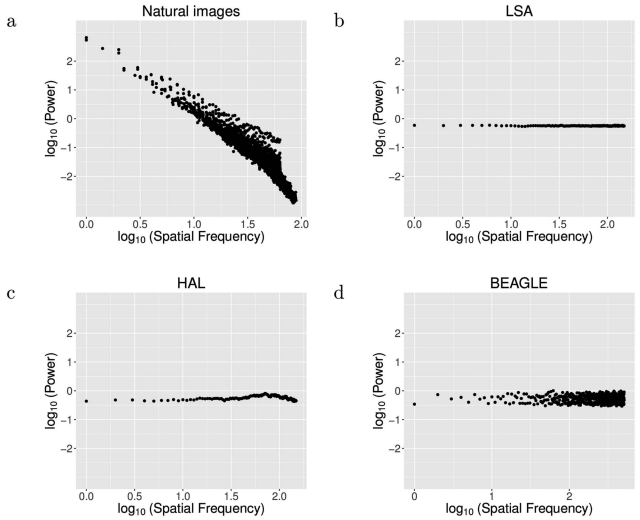


Figure 16. Power spectra of item representations for three semantic-memory models, LSA (b), HAL (c), and BEAGLE (d), compared with natural images (a). See main text for details. Natural images have the usual coloured-noise spectrum reflecting auto-correlation in across vector dimensions, whereas vectors in the three semantic models are uncorrelated, approximating white noise.

Although it is possible that decorrelated representations are used by the brain to satisfy the statistical properties necessary to use convolution, it also could be the reverse: that information is already decorrelated in the brain for other "reasons," and under those conditions, convolution becomes an effective way to store and retrieve associations.

Finally, this work builds on the work of (Kelly, 2010; Kelly et al., 2013) in the following ways. First, we demonstrated that Kelly et al.'s permutation representation succeeds in additional conditions: heteroassociations, and multiple associations stored, extending Kelly and colleagues' demonstrations involving a single auto-association. Second, we address the question of neural plausibility of various representation types. Third, we introduced the idea of DoG-transformed natural images as an alternative potential solution to the same problem as Kelly and colleagues were concerned with.

In summary, our demonstrations show that the way in which the brain suppresses auto-correlations, normally a nuisance for convolution, produces conditions that are compatible with convolution-based association-memory.

Résumé

La convolution est une opération mathématique utilisée dans les modèles-vecteurs de la mémoire s'étant avérés fructueux pour

expliquer une vaste gamme de comportements, dont la mémoire par associations entre paires d'items, une prémisse importante de la mémoire à partir de laquelle une grande variété de comportements de mémoire de tous les jours dépendent. Or, les modèles de convolution ont de la difficulté avec la représentation d'items naturalistes, lesquels sont hautement auto-corrélés (tel qu'on peut voir par ex. sur des photographies), et cela remettrait en cause leur plausibilité neurale. Par conséquent, les modeleurs travaillant avec la convolution ont utilisé les représentations d'items composées de valeurs aléatoires, mais l'introduction de représentations s'apparentant au bruit soulève la question à savoir comment ces valeurs d'apparence aléatoire pourraient être reliées aux réelles propriétés d'item. Nous proposons qu'il existe probablement déjà une solution de compromis à ce problème. Il est aussi connu depuis longtemps que le cerveau tend à réduire les auto-corrélations au niveau de ses entrées. Par exemple, les cellules du centre-périphérie de la rétine se rapprochent d'une différence de gaussiennes. Cela avantage la périphérie mais transforme aussi les images naturelles en images ressemblant à ce qu'on qualifie statistiquement de bruit de fond. Nous montrons comment les images transformées par différences de gaussiennes, même si non optimales en comparaison aux représentations s'apparentant au bruit, survivent davantage au modèle de convolution que les images naturalistes. Ceci est démonstration du principe que la tendance généralisée du cerveau à réduire les auto-corrélations pourrait résulter de représentations de l'information étant déjà adéquatement compatibles avec la convolution, venant appuyer la plausibilité neurale de la mémoire associative basée sur la convolution.

Mots-clés : convolution, mémoire associative, rappel indicé, représentations, modèles mathématiques.

References

Anderson, J. A. (1970). Two models for memory organization using interacting traces. *Mathematical Biosciences, 8,* 137–160.

Balboa, R. M., & Grzywacz, N. M. (2000). The role of early retinal lateral inhibition: More than maximizing luminance information. *Visual Neuroscience, 17,* 77–89.

Burgess, C. (1998). From simple associations to the building blocks of language: Modeling meaning in memory with the hal model. *Behavior Research Methods, Instruments, and Computers, 30,* 188–198.

Caplan, J. B. (2011). *Grid cells and place cells as proof-of-principle for convolution-based association-memory in the medial temporal lobe*. International Conference on Memory, York, United Kingdom.

Dong, D. W., & Atick, J. J. (1995). Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network: Computation in Neural Systems, 6,* 159–178.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Optical Society of America, 4,* 2379–2394.

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., & Moser, M. (2004). Spatial representation in the entorhinal cortex. *Science, 305,* 1258.

Günther, F. (2015). *Repository for semantic spaces*. Retrieved from http://www.lingexp.uni-tuebingen.de/z2/LSAspaces/

Hafting, T., Fyhn, M., Molden, S., Moser, M-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature, 436,* 801–806.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review, 96,* 208–233.

Jones, N. M., & Mewhort, J. K. D. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review, 114,* 1–37.

Kahana, M. J. (2012). *Foundations of human memory*. New York, NY: Oxford University Press.

Kelly, M. A. (2010). Advancing the theory and utility of holographic reduced representations (Master's thesis, Queen's University, Canada). *ProQuest Dissertations and Theses*. Retrieved from http://search.proquest.com/docview/853293422

Kelly, M. A., Blostein, D., & Mewhort, D. J. K. (2013). Encoding structure in holographic reduced representations. *Canadian Journal of Experimental Psychology, 67,* 79–93.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240.

Longuet-Higgins, H. C. (1968). Holographic model of temporal recall. *Nature, 217,* 104.

Luce, R. D. (1959). *Individual choice behavior*. New York, NY: Wiley.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers, 28,* 203–208.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Biological Sciences, 207,* 187–217.

Metcalfe Eich, J. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Murdock, B. B. (1985). Convolution and matrix systems: A reply to Pike. *Psychological Review, 92,* 130–132.

Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review, 91,* 281–294.

Plate, T. A. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks, 6,* 623–641.

Plate, T. A. (2003). *Holographic reduced representation*. Stanford, CA: CSLI Publication.

Renart, A., Rocha, J. de la, Bartho, P., Hollender, L., Parga, N., Reyes, A., & Harris, K. D. (2010). The asynchronous state in cortical circuits. *Science, 327,* 587–590.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323,* 533–536.

Solstad, T., Moser, E. I., & Einevoll, G. T. (2006). From grid cells to place cells: A mathematical model. *Hippocampus, 16,* 1026–1031.

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London Series B: Biological Sciences, 216,* 427–459. http://dx.doi.org/10.1098/rspb.1982.0085

Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science, 262,* 685–688.

Tanaka, K. (2003). Columns for complex visual object features in the inferotemporal cortex: Clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex, 13,* 90–99.

Wang, G., Tanaka, K., & Tanifuji, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. *Science, 272,* 1665–1668.

Received October 30, 2015
Accepted November 3, 2016 ■