

REVIEW

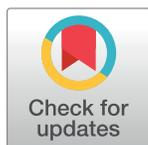
Neural correlates of sparse coding and dimensionality reduction

Michael Beyeler^{1,2,3,4}*, Emily L. Rounds⁵, Kristofor D. Carlson^{5,6}, Nikil Dutt^{4,5}, Jeffrey L. Krichmar^{4,5}

1 Department of Psychology, University of Washington, Seattle, Washington, United States of America, **2** Institute for Neuroengineering, University of Washington, Seattle, Washington, United States of America, **3** eScience Institute, University of Washington, Seattle, Washington, United States of America, **4** Department of Computer Science, University of California, Irvine, California, United States of America, **5** Department of Cognitive Sciences, University of California, Irvine, California, United States of America, **6** Sandia National Laboratories, Albuquerque, New Mexico, United States of America

* These authors contributed equally to this work.

* mbeyeler@uw.edu



OPEN ACCESS

Citation: Beyeler M, Rounds EL, Carlson KD, Dutt N, Krichmar JL (2019) Neural correlates of sparse coding and dimensionality reduction. *PLoS Comput Biol* 15(6): e1006908. <https://doi.org/10.1371/journal.pcbi.1006908>

Editor: Aldo A. Faisal, Imperial College London, UNITED KINGDOM

Published: June 27, 2019

Copyright: © 2019 Beyeler et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: MB, ND, and JLK were supported by the National Science Foundation (Award IIS-1302125). In addition, MB was supported by the Washington Research Foundation Funds for Innovation in Neuroengineering and Data-Intensive Discovery. JLK was supported in part by the Defense Advanced Research Projects Agency (DARPA) via Air Force Research Laboratory (AFRL) Contract No. FA8750-18-C-0103 (Lifelong Learning Machines: L2M). Sandia National Laboratories is a multiprogram laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. The funders had no role in study design, data collection and analysis,

Abstract

Supported by recent computational studies, there is increasing evidence that a wide range of neuronal responses can be understood as an emergent property of nonnegative sparse coding (NSC), an efficient population coding scheme based on dimensionality reduction and sparsity constraints. We review evidence that NSC might be employed by sensory areas to efficiently encode external stimulus spaces, by some associative areas to conjunctively represent multiple behaviorally relevant variables, and possibly by the basal ganglia to coordinate movement. In addition, NSC might provide a useful theoretical framework under which to understand the often complex and nonintuitive response properties of neurons in other brain areas. Although NSC might not apply to all brain areas (for example, motor or executive function areas) the success of NSC-based models, especially in sensory areas, warrants further investigation for neural correlates in other regions.

Author summary

Brains face the fundamental challenge of extracting relevant information from high-dimensional external stimuli in order to form the neural basis that can guide an organism's behavior and its interaction with the world. One potential approach to addressing this challenge is to reduce the number of variables required to represent a particular input space (i.e., dimensionality reduction). We review compelling evidence that a range of neuronal responses can be understood as an emergent property of nonnegative sparse coding (NSC)—a form of efficient population coding due to dimensionality reduction and sparsity constraints.

Introduction

Brains face the fundamental challenge of extracting relevant information from high-dimensional external stimuli in order to form the neural basis that can guide an organism's behavior

decision to publish, or preparation of the manuscript. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA, AFRL, or the US Government.

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

and its interaction with the world. To support complex patterns of behavior, populations of interconnected neurons must implement a rich repertoire of linear and nonlinear operations on their synaptic inputs that take into account context, experience, and anatomical constraints [1]. For example, anatomical bottlenecks often force the information stored in a large number of neurons to be compressed into an orders-of-magnitude-smaller population of downstream neurons [2–4], such as storing information from 100 million photoreceptors in 1 million optic nerve fibers or resulting in a 10–10,000-fold convergence from cortex to the basal ganglia [3].

One potential approach to addressing this challenge is to reduce the number of signals required to transmit information in the network—for example, through **sparse-coding** schemes (text in bold appear in the Glossary section), in which information is represented by the activity of a small proportion of neurons in a population [5–7]. A number of different definitions of sparsity can be found in the literature [8, 9], which can sometimes lead to controversy as to which codes can still be considered sparse [8]. An extreme example is the so-called local code, in which each unique event, or “context,” is encoded by a single active neuron, or “grandmother cell” [10] (illustrated in the left column of Fig 1A). Local codes not only suffer from low **representational capacity**, because they allow a population of N neurons to encode at most N contexts, but also require a large number of neurons to cover the space of possible contexts. On the other hand, a dense code represents each context by the combined activity of all neurons in the population (Fig 1A, right column). In theory, dense codes lead to high representational capacity (at M activity levels, allowing for M^N contexts to be encoded), but they also suffer from neuronal cross talk because every neuron is involved in every context. Alternatively, sparse codes (Fig 1A, center column) can be described as a trade-off between the benefits and drawbacks of dense and local codes, in which each context is encoded by a different subset of neurons in the population. [5]. In general, sparse coding reduces the overall neural activity necessary to represent information.

Another approach to address this challenge is to reduce the number of variables required to represent a particular input, stimulus, or task space, a process known as **dimensionality reduction**. Although responses of individual neurons are often complex and highly nonlinear, a

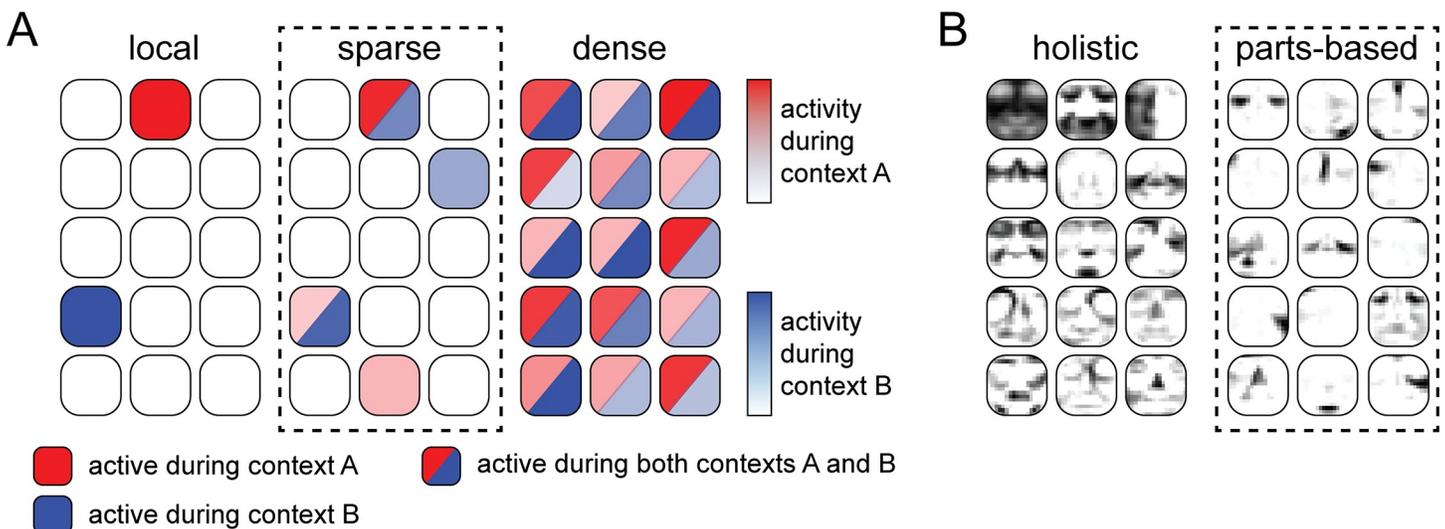


Fig 1. NSC promotes population codes that are both sparse and parts based. (A) Hypothetical activity in a population of neurons during presentation of two different external stimuli (“contexts”). A sparse code is a trade-off between a local code (in which a context is represented by the activity of a single neuron, and different contexts are represented by different neurons) and a dense code (in which all neurons are active, and their combined activity is used to encode each context). Dense codes possess great memory capacity but suffer from cross talk among neurons, whereas local codes do not suffer from interference but also have no capacity for generalization (inspired by [8]). (B) In a holistic representation of faces, individual neurons in the population respond themselves to faces as a whole [11], whereas in a parts-based representation, individual neurons explicitly encode individual face components [12], such as the eyes, nose, and mouth (inspired by [13]). NSC, nonnegative sparse coding.

<https://doi.org/10.1371/journal.pcbi.1006908.g001>

population of neurons might share activity patterns because of individual neurons in the population not being independent of each other. Dimensionality reduction methods have proved useful in elucidating these shared activity patterns and thus effectively explaining population activity using a lower number of variables than there are neurons in the population (for a recent review, see [14]).

Neurons often encode several behaviorally relevant variables simultaneously [15–18], allowing for multifaceted representations of high-dimensional stimulus spaces. For example, a population of neurons tasked with encoding human faces might opt to represent each individual face as a combination of a set of standard faces (Fig 1B, left column). In such a **holistic representation** of faces [11], each individual neuron would itself respond to a face as a whole (i.e., a face “template”) without explicitly representing individual face components, and an arbitrary face could be represented by combining different face templates (e.g., by adding 10% of template 1 to 20% of template 2 and subtracting 30% of template 3). On the other hand, faces can also be represented as a combination of individual face components, such as eyes, noses, and mouth, in what is known as a **parts-based representation** (Fig 1B, right column) [12, 19]. Both approaches allow for representing arbitrary faces as a combination of neural activity but have drastically different consequences on the set of stimulus features each neuron responds to. Although visual information from the eyes, nose, and mouth would of course be included in a holistic face representation, that information would not be explicitly represented as structural units in their own right [11]. Linear combinations of holistic components often involve complex cancellations between positive and negative contributions and thus lack the intuitive meaning of adding parts to form a whole. In contrast, a parts-based representation allows for only nonsubtractive combinations of stimulus features [12]. Although the relevant stimulus dimensions are often not known a priori, several sophisticated mathematical techniques exist that allow us to discover these representations directly from experimental data [14, 19–23].

In this article, we review evidence from experimental and theoretical studies suggesting that a number of neuronal responses can be understood as an emergent property of nonnegative sparse coding (NSC), an efficient population coding scheme based on dimensionality reduction and sparsity constraints. In particular, we review evidence for NSC in sensory areas that efficiently encode external stimulus spaces, for associative areas to conjunctively represent multiple behaviorally relevant variables, and for the basal ganglia to coordinate movement.

Nonnegative sparse coding as a modern variant of the efficient coding hypothesis

Efficient coding

The fundamental principle of **efficient coding** is that a sensory system is adjusted to the specific statistics of the natural environment from which it encodes and transmits information [24–27]. Efficiency, in this context, is an information-theoretic term that should not be confused with “minimizing energy expenditure.” Instead, a sensory pathway is treated as a noisy communication channel, in which the goal is to maximize the rate at which information can be reliably transmitted by minimizing the redundancy between representational units.

Early theories of efficient coding [24, 25] were developed based on the visual system. Attneave [25] pointed out that there is a significant degree of redundancy in natural visual images because of correlations in both the spatial and temporal domains (for a recent review, see [28]). For example, the luminance values of a pair of pixels separated by a fixed distance in a natural image are likely to be highly correlated (Fig 2A). These statistical regularities constrain the images a visual system is likely to encounter to a tiny fraction of the set of all possible images. It was therefore argued that the visual system should not waste resources on processing

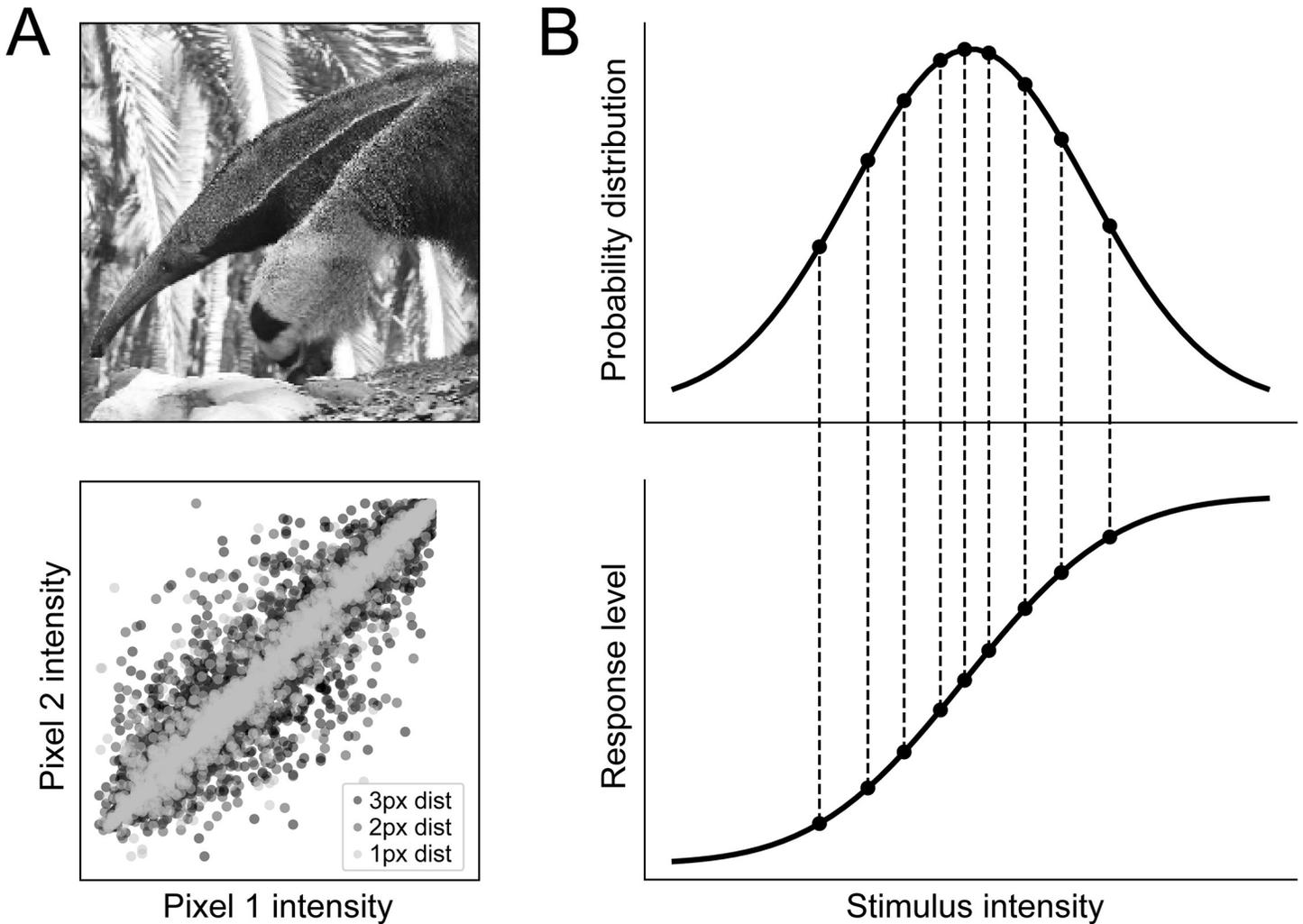


Fig 2. Efficient coding hypothesis. (A) Sensory stimuli in the environment, such as an image of an anteater, display significant statistical structure. For example, the luminance value of nearby pixels in the image is significantly correlated, an effect that exists even for nonadjacent pixels (inspired by [27]). Neural systems can improve their coding efficiency by accounting for and reducing such information redundancy. (B) For a given distribution of sensory characteristics in the world (top), a neuron's information capacity is maximized when all response levels are used with equal frequency (inspired by [29]). Intervals between each response level encompass an equal area under the intensity distribution, so each state is used with equal frequency.

<https://doi.org/10.1371/journal.pcbi.1006908.g002>

arbitrary images but instead use statistical knowledge about its environment to represent the relevant input space as economically as possible.

Extending this idea to the neural, Barlow [24] proposed that the goal of early neurons in sensory processing is to transform raw visual inputs into an efficient representation such that as much information as possible can be extracted from them given limited neural resources. This efficient coding principle has been able to explain a wide variety of neuronal response properties in the early visual system, such as the center-surround structure of receptive fields (RFs) in the retina [30], temporally decorrelated signals in the lateral geniculate nucleus (LGN) [31], and the coding of natural scenes in the primary visual cortex (V1) [9].

At the level of single neurons, efficient coding suggests that the information carried by a neuron's response can be maximized by using all response levels with equal frequency [29, 32, 33]. For example, in the case of a neuron representing a single input variable with a single output variable, information is maximized when the input–output function corresponds to the

cumulative probability function for the different input levels [29], as shown in Fig 2B. Note that this coding procedure amplifies inputs in proportion to their expected frequency of occurrence rather than reserving large portions of its dynamic range for improbable inputs [29, 32]. On the other hand, if the input–output function sensitivity is chosen as too low, high levels of the stimulus feature will be indistinguishable as the response function saturates; if the sensitivity is set too high, low levels of the stimulus feature cannot drive responses [29].

At the level of neuronal populations, neural responses should be both decorrelated (i.e., independent from one another) and sparse (i.e., involve only a small fraction of neurons in the population) [27].

Sparse coding

Taking these ideas a step further, Olshausen and Field [34] noted that natural images contain statistical dependencies beyond linear pairwise correlations among image pixels and argued that these higher-order correlations should be taken into account when developing an efficient code. Their goal was thus to find a linear coding strategy capable of reducing these higher-order forms of redundancy.

Linear sparse coding is one such strategy, in which monochromatic images $I(x,y)$ are described in terms of a linear superposition of a number of B **basis functions**, $w_b(x,y)$:

$$I(x,y) = \sum_{b=1}^B w_b(x,y)h_b, \quad (1)$$

where h_b are stochastic coefficients that are different for each image [35, 36]. Learning a sparse code for images thus involved determining the values of both $w_b(x,y)$ and h_b for all b and (x,y) , given a sufficient number of observation of images, under the constraint that h_b be sparse. In this context, h_b was considered sparse if it took very small or very large (absolute) values more often than a Gaussian random variable would [36]. This sparsity constraint allowed for basis functions that were not needed to describe a given image structure to be weeded out.

When Olshausen and Field applied linear sparse coding to natural images, they found that the emerging basis functions were qualitatively similar in form to RFs of simple cells in V1 [35, 37], thus giving empirically observed RFs an information-theoretic explanation. In this context, h_b in Eq 1 corresponded to the (signed) activation value of a particular V1 neuron, and $w_b(x,y)$ were the connection weights (or synaptic weights in an artificial neural network) that were closely related to that neuron’s RF.

Sparsity, in this context, is an information-theoretic concept related to how efficiently and completely information is encoded with the basis functions described previously. Please note that this is different from empirical observations of brain areas being “sparsely” activated; that is, sparse population activity does not necessarily imply that a brain area implements a sparse-coding scheme. This confusion is fueled in part by the wide variety of definitions of sparsity used in the literature [8, 38]. For example, even though sparse coding (as a theoretical framework) applied to natural images yields V1-like RFs, recent evidence suggests that neural activity in V1 might not be as sparsely activated as previously thought [39, 40]. However, V1 still codes stimuli efficiently [40].

Olshausen and Field went on to show that the set of basis functions that best described V1 RFs was greater in number than the effective dimensionality of the input (which they termed an overcomplete basis set) [37]. It is worth noting that sparse coding with an overcomplete basis set is typically associated with an anatomical fan-out motif, such as expanding 1 million optic nerve fibers into more than 100 million V1 neurons or from a small number of mossy fibers to a 100-fold-larger number of granule cells in the cerebellum.

However, as pointed out by Hoyer [41], linear sparse coding falls short of providing a literal interpretation for V1 simple-cell behavior for two reasons: (1) every neuron could be either positively or negatively active, and (2) the input to the neural network was typically double-signed, whereas V1 neurons receive visual input from the LGN in the form of separated, non-negative ON and OFF channels.

In order to transform Olshausen and Field's sparse coding from a relatively abstract model of image representation into a biologically plausible model of early visual cortex processing, Hoyer [41, 42] thus proposed to enforce both input signal and neuronal activation to be non-negative (though still allowing inhibitory connections). This seemingly simple change had remarkable consequences on the quality of the sensory representation: whereas elementary image features in the standard sparse-coding model could "cancel each other out" through subtractive interactions, enforcing nonnegativity ensured that features combined additively, much like the intuitive notion of combining parts to form a whole. The resulting parts-based representations resembled RFs in V1 much more closely than other holistic representations. These considerations led to the formulation of NSC in its current form.

Nonnegative sparse coding

As a special case of linear sparse coding, NSC shares the same goal of accurately describing observed data as a superposition of a set of sparsely activated basis functions, as well as enforcing dimensionality reduction. In addition, NSC requires all basis functions and activation values (i.e., $w_b(x,y)$ and h_b in Eq 1) to be nonnegative. However, NSC is more than just linear sparse coding with nonnegative weights. For example, whereas linear sparse coding typically uses a larger number of basis functions than there are dimensions in the input (thus achieving dimensionality expansion), NSC makes use of **nonnegative matrix factorization (NMF)** to achieve dimensionality reduction. This has interesting implications for the kinds of basis functions that can be learned. Most prominently, the nonnegativity constraints used in NMF force the different basis functions to add up linearly, thus leading to the distinctive parts-based representations.

Consider S observed stimuli or data samples, each composed of F observed feature values, such as a collection of S images $I(x,y)_s$ ($s \in [1, \dots, S]$) from the previous example, each consisting of F different grayscale values. If we arrange the observed feature values of the s -th observation into a vector \vec{v}_s (i.e., by flattening each observed image), and if we arrange all vectors into the columns of an $F \times S$ data matrix \mathbf{V} , then linear decompositions describe these data as

$$\mathbf{V} \approx \mathbf{WH}, \tag{2}$$

where \mathbf{W} is an $F \times B$ matrix that contains as its columns the B basis functions of the decomposition (i.e., the b -th column of \mathbf{W} corresponding to $w_b(x,y) \forall x,y$ in Eq 1), and \mathbf{H} is a $B \times S$ matrix containing as its columns the activation values of each basis function for a particular input stimulus (i.e., the b -th column of \mathbf{H} corresponding to $h_b \forall b$ in Eq 1). The difference between \mathbf{V} and \mathbf{WH} is termed the reconstruction error.

The goal of NSC is then to find a linear decomposition of \mathbf{V} that minimizes the reconstruction error while guaranteeing that \mathbf{H} is sparse. This can be achieved by minimizing the following cost function [42]:

$$\min_{\mathbf{W}, \mathbf{H}} \frac{1}{2} \|\mathbf{V} - \mathbf{WH}\|^2 + \lambda \sum_{ij} f(\mathbf{H}_{ij}), \tag{3}$$

subject to the constraints $\forall ij: \mathbf{W}_{ij} \geq 0, \mathbf{H}_{ij} \geq 0$, and $\|\vec{w}_i\| = 1$, where \vec{w}_i denotes the i -th column of \mathbf{W} . Here, the left-hand term describes the reconstruction error, whereas the right-hand

term describes the sparsity of the decomposition. The trade-off between accurate reconstruction and sparsity is controlled by the parameter λ (where $\lambda \geq 0$), whereas the form of f defines how sparsity is measured (a typical choice is the L1 norm on \mathbf{H}).

Analogous to efficient coding, Eq 3 forces prediction errors to be amplified in proportion to their expected frequency of occurrence because a more frequent event would show up more frequently in \mathbf{V} . Hence, accounting for a rare observation at the expense of ignoring a more common one would result in an increased reconstruction error.

In the case of $\lambda = 0$, Eq 3 reduces to the squared-error version of NMF. Although NMF enforces all elements of \mathbf{W} and \mathbf{H} to be nonnegative, the resulting decomposition might not be sparse, depending on the number of basis functions B . In order to emphasize decompositions in which \mathbf{H} is sparse, Eq 3 should be minimized with $\lambda > 0$ [42].

Another open parameter is the number of basis functions, B , which controls the predictive power of the model and must be determined empirically. With a small number of basis functions, NSC is unlikely to achieve a low reconstruction error, be it in familiar contexts (training data) or in novel contexts (held-out test data). In this case, the error depends on the systematic bias of the model, and the model is said to underfit the data (left-hand side of Fig 3). With increased model complexity, the model can learn subtle differences between different contexts with high accuracy, leading to a reduced bias (training) error. However, with increased complexity, the model is more likely to learn patterns between training contexts that arise either from underlying noise or from spurious correlations. As a result, the model will respond according to these learned patterns when a novel context is presented (rather than according to the underlying actual relationships), in which case the model is said to overfit the data (right-hand side of Fig 3). Hence, the goal of a successful model is to find the ideal compromise in the bias–variance error trade-off [43] (labeled “best model” in Fig 3).

Analogously to [35, 37], the basis functions obtained in NSC can be interpreted as the connection weights of a population of simulated neurons in an artificial neural network. In other words, under NSC, the number of basis functions B corresponds to the number of output neurons, and the response of the b -th model output neuron ($b \in [1, \dots, B]$) to a particular input stimulus s , termed r_{bs} , can be computed by feeding the dot product of that neuron’s connection weights (i.e., the b -th column in \mathbf{W} , \vec{w}_b) and a data vector (i.e., the s -th column in \mathbf{V} , \vec{v}_s) to an activation function Θ :

$$r_{bs} = \Theta(\vec{w}_b \cdot \vec{v}_s), \tag{4}$$

where “ \cdot ” denotes the dot product. For example, the linear response of a model neuron can be calculated by setting Θ to the identity function $\Theta(x) = x$. Note that the response of the model neuron to different stimuli $s \in [1, \dots, S]$ involves different columns of \mathbf{V} but always relies on \vec{w}_b .

Thus, we can utilize \mathbf{W} (which must remain fixed once learned) and Eq 4 to simulate a model neuron’s response to arbitrary input stimuli by replacing the column in \mathbf{V} with new input. This allows us to investigate the response properties of individual model neurons much in the same way that experimental neuroscientists study biological neurons. This is important because it means that NSC can be used to model neural activity in the brain, and the resulting activity patterns generated by NSC can be compared to and evaluated against experimental findings.

It is important to note that the absence of negative weights in Eqs 2–4 does not preclude the modeling of inhibitory connections or even posit that inhibitory connections cannot participate in NSC. Rather, one important aspect of NSC is the parts-based, NMF-like decomposition of \mathbf{V} ; one way to achieve this is by enforcing nonnegativity constraints on \mathbf{W} and \mathbf{H} . Several

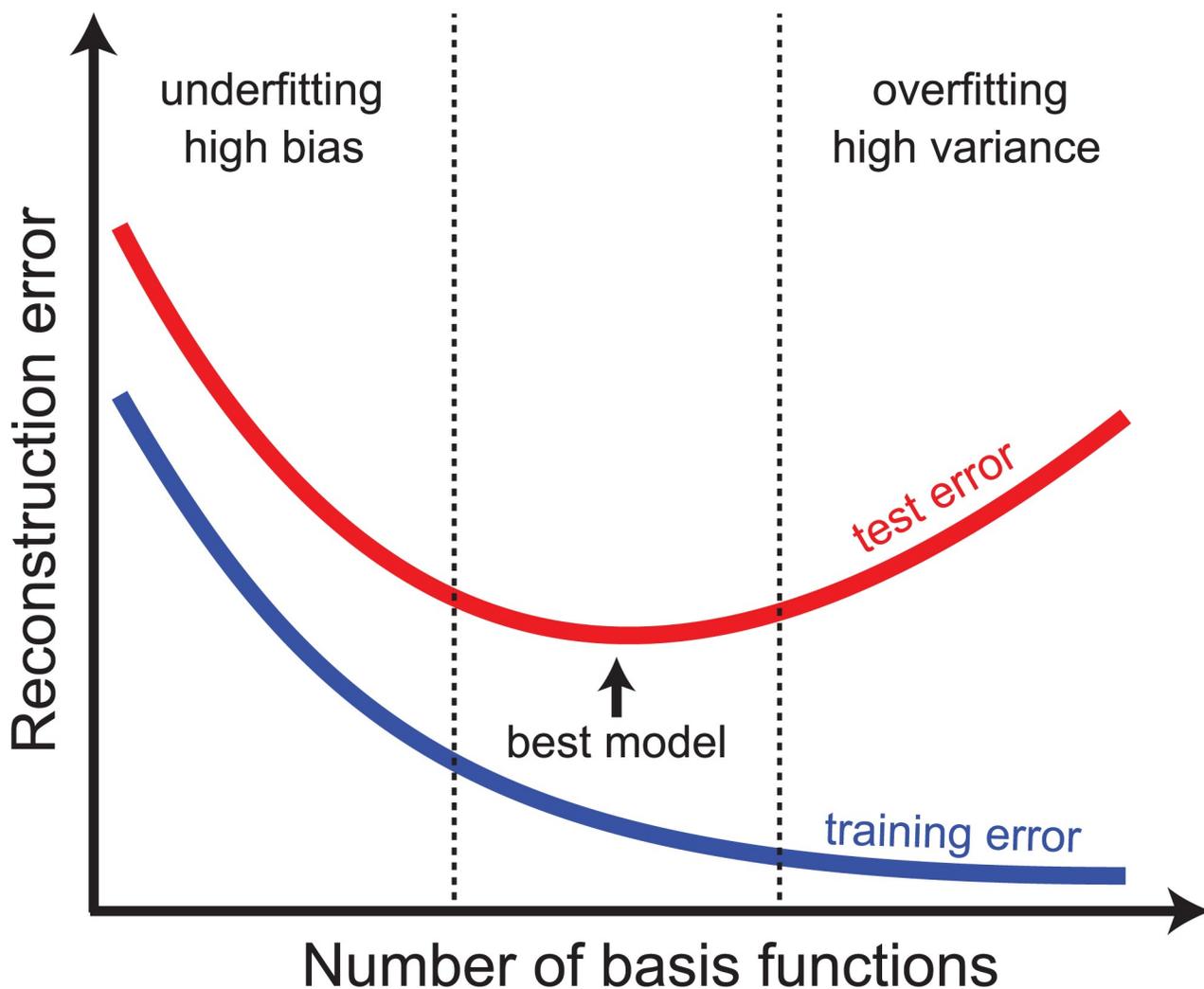


Fig 3. The bias–variance dilemma. With increased model complexity (i.e., with an increased number of basis functions), the reconstruction error on a set of familiar (training) data typically decreases until it reaches zero. In contrast, the reconstruction error on a set of unfamiliar, held-out (test) data typically goes through a minimum as a function of model complexity. A successful model chooses the number of basis functions such that the generalization (test) error is minimized (labeled “best model”).

<https://doi.org/10.1371/journal.pcbi.1006908.g003>

studies have successfully incorporated inhibitory connections into their NSC-based models. One approach is to model them as nonnegative synaptic conductances. For example, Hoyer [41] used NSC to model V1 neurons as receiving input from both excitatory ON and inhibitory OFF cells in the LGN. Using prewhitened natural images, Hoyer sampled 12×12 pixel patches from the images and then separated positive and negative values into separate channels. Each image patch was thus represented by a $2 \times 12 \times 12 = 288$ dimensional vector, each element of which mimicked the activity of an ON or OFF cell in response to the image patch. These vectors were then arranged into the columns of \mathbf{V} . This procedure not only preserved the parts-based quality of the encoding but also allowed the modeling of the convergence of ON and OFF pathways. Another approach is to drop the nonnegativity constraint on \mathbf{W} and thus effectively operate with both positive and negative synaptic weights. Only recently did it become clear that this approach was able to preserve the parts-based quality of the encoding (as long as nonnegativity of \mathbf{H} was enforced) [44], thus simplifying the construction of more complex network topologies.

Empirical evidence for NSC in the brain

In this section, we review evidence for NSC in several brain regions. In particular, NSC has been observed in sensory areas, an association cortex area, and the basal ganglia. Although these findings suggest that NSC might apply elsewhere in the brain, thus warranting further investigation, we are aware that NSC does not apply to everywhere in the brain. We will further discuss the limits of NSC in the Model limitations section of the Discussion.

Because of its roots in efficient coding theories of natural image processing, there is a large body of research highlighting the role of NSC in visual cortex function (e.g., [24, 35, 41, 45, 46]). More recently, NSC-like computational models have found application outside visual cortex, where they have started to provide compelling evidence that a wide variety of neuronal response properties might be understood as an epiphenomenon of efficient population coding based on dimensionality reduction. Examples include elucidating the dimensions along which perceptual space is organized in the olfactory system [47, 48], the coordination of movement in the cortico-basal ganglia-thalamo-cortical loop [3, 49], and the combined representation of allocentric and route-based spatial navigation cues in retrosplenial cortex (RSC) [50].

In the following subsections, we review studies describing evidence for NSC that either successfully explains response properties of individual neurons or has been instrumental in elucidating the dynamics at the population level. We start this section with some early modeling work that shows parts-based dimensionality reduction analogous to neuronal responses in inferotemporal cortex (IT).

NSC in the inferotemporal cortex

The notion of parts-based object recognition is compatible with hierarchical models of vision, in which activation of simple features feeds into the activation of complex features [51]. There is a long history of debate as to whether humans detect faces based on their individual parts or as correctly arranged wholes (for reviews, see [11, 52, 53]). The working hypothesis is that the brain might use holistic face information as an early gating mechanism to allow visual stimuli access to the face processing module but that most cortical circuitry relies on parts-based information [53]. Converging evidence from human imaging studies and primate physiology suggests that faces are processed in localized “patches” within IT [54], where cells detect distinct constellations of face parts [55, 56], such as eyes [57], and that whole faces can be recognized by taking linear combinations of neuronal activity across IT [19, 58].

An influential paper by Lee and Seung [13] found that applying NMF to a database of face images yielded sparse, localized features that resembled parts of a face (Fig 4A) in a similar fashion to responses in area IT. In their case, NMF acted on an $F \times S$ data matrix \mathbf{V} , whose rows corresponded to distinct features of the input (e.g., F different pixels of an image) and whose columns corresponded to different stimuli or observations of those features (e.g., S different images). NMF was used to decompose the matrix into two reduced-rank matrices (Fig 4, inset) whose linear combination could be weighted such that the product of \mathbf{W} and \mathbf{H} provided an accurate reconstruction of \mathbf{V} (see Eq 2).

A particular image, in this case encoded by $F = 19 \times 19 = 361$ pixels could be accurately represented by a linear combination of a small number ($B = 49$) of encoding variables or “basis images” (Fig 4A). Such a representation is reminiscent to neural processing in IT, an area in the ventral visual “what” stream involved in encoding high-level object identity [58, 59], in which images of whole faces can be linearly reconstructed using responses of approximately 200 neurons that each respond to a certain set of physical facial features [19].

Interestingly, such a parts-based representation is not specific to face processing in IT; the same principle can be extended to body-selective regions in IT [60, 61].

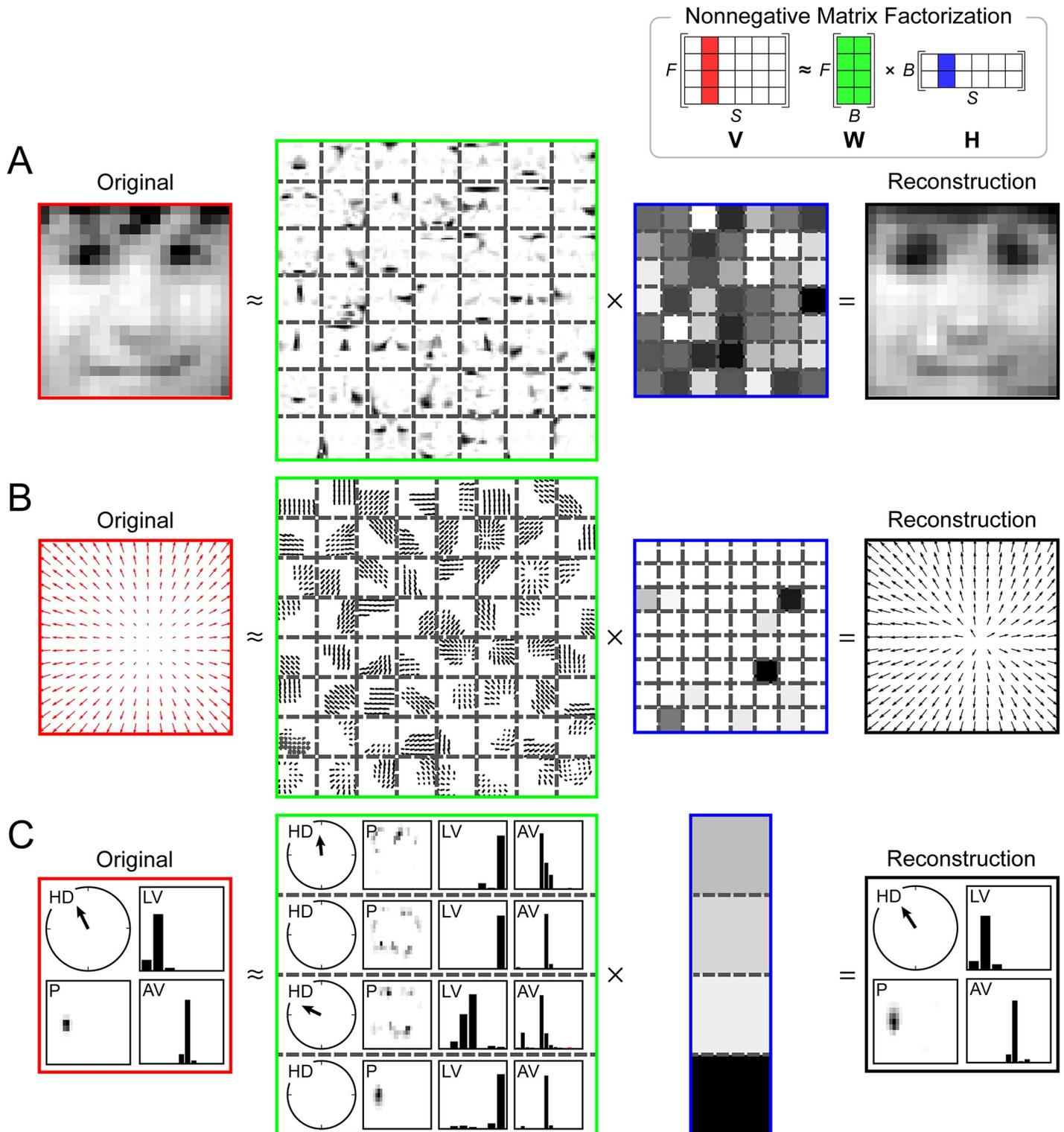


Fig 4. Sparse and parts-based representations recovered by NMF resemble RFs across brain regions. NMF (inset) can reconstruct a data matrix V (F features \times S stimuli) from two reduced-rank matrices W (containing B basis functions) and H (containing the hidden coefficients of the decomposition). Any individual input stimulus (i.e., column in V , red) can be reconstructed from a linear combination (i.e., column in H , blue) of a set of basis functions (i.e., all columns in W , green). (A) A facial image can be reconstructed from a sparse activation of simulated IT neurons that preferentially respond to parts of faces (inspired by [13]). (B) An optic flow field can be reconstructed from a sparse activation of model MSTd neurons that prefer various directions of 3D self-translation and self-rotation. (C) A rat's 2D allocentric position and route-based direction of motion can be reconstructed from a sparse activation of model RSC neurons that prefer an intricate combination of LV, AV, HD,

and P. For the sake of clarity, only the four most contributing hidden coefficients (out of 30) are shown. AV, angular velocity; HD, head direction; IT, inferotemporal cortex; LV, linear velocity; MSTd, dorsal subregion of the medial superior temporal area; NMF, nonnegative matrix factorization; P, 2D position; RSC, retrosplenial cortex. *Adapted with permission from [46].*

<https://doi.org/10.1371/journal.pcbi.1006908.g004>

Although there seems to be a consensus that information-theoretic explanations are relevant when investigating early sensory areas, higher-order brain areas are often considered to be specialized for performing tasks (e.g., recognizing objects, making decisions, navigating an environment) rather than the efficient encoding of information. It is therefore possible that the essential components of NSC might well be present in higher-order areas but, to date, have gone unnoticed.

NSC in the retina

Because of its roots in efficient coding theories of natural image processing, NSC figures prominently in the vision neuroscience literature. For example, NMF-based models were able to reconstruct in vitro neuronal spike trains from the salamander retina [44, 62]. By combining spike-triggered average with NMF, Liu and colleagues [44] were able to identify the subunit layout of retinal ganglion cells (Fig 5). This technique, termed spike-triggered NMF (STNMF), involved applying NMF to the collection of those stimulus patterns contained in a spatiotemporal white-noise sequence that caused a given neuron to spike. Akin to common reverse-correlation analysis, the researchers averaged the collection of spike-eliciting stimulus segments to form the spike-triggered stimulus ensemble (Fig 5A). STNMF then decomposed the ensemble of effective spike-triggered stimuli into a matrix \mathbf{W} containing a set of modules (or basis functions) and a matrix \mathbf{H} containing a set of hidden coefficients.

Intuitively, the modules derived by STNMF should capture the subunit decomposition of the cell's RF because the spike-eliciting stimuli should have essential statistical structure imprinted on them by the subunits, such as correlations between pixel values [44]. And indeed, the identified modules corresponded to individual presynaptic bipolar cells, as verified by multi-electrode array recordings with simultaneous recordings from individual bipolar cells through sharp microelectrodes [44]. This allowed the researchers to improve predictions about how ganglion cells respond to natural stimuli without the need to guess a specific model structure that may be constrained in terms of the size, shape, number, or nonlinearity of ganglion cell subunits.

NSC in the early visual cortex

NSC has been extensively applied to early visual cortex, where it has successfully explained orientation and frequency tuning of simple and complex cells in V1 [41] as well as edge-like pooling of spatial frequency channels in V2 [63], including RF properties such as end-stopping and contour integration [64]. These theoretical findings are in good agreement with a large body of research documenting the sensory response of V1 across animal models (e.g., [65–68]), although they are not without controversy. For example, one study [67] criticized that some of the early sparse-coding models generated RFs that looked like stereotyped edge detectors and did not capture the diversity of RF structure observed in cat and monkey V1. However, by adjusting these models to limit the number of active neurons (“hard” sparsity) instead of limiting mean neuronal activity (“soft” sparsity), Rehn and Sommer [69] were able to account for the diversity of shapes in biological RFs. Other researchers were concerned that the apparent sparse activation of V1 was an artifact of using simple artificial stimuli such as sinusoidal gratings and drifting bars, but Vinje and Gallant [9] were able to show that natural viewing conditions actually increased the sparsity of V1 activation.

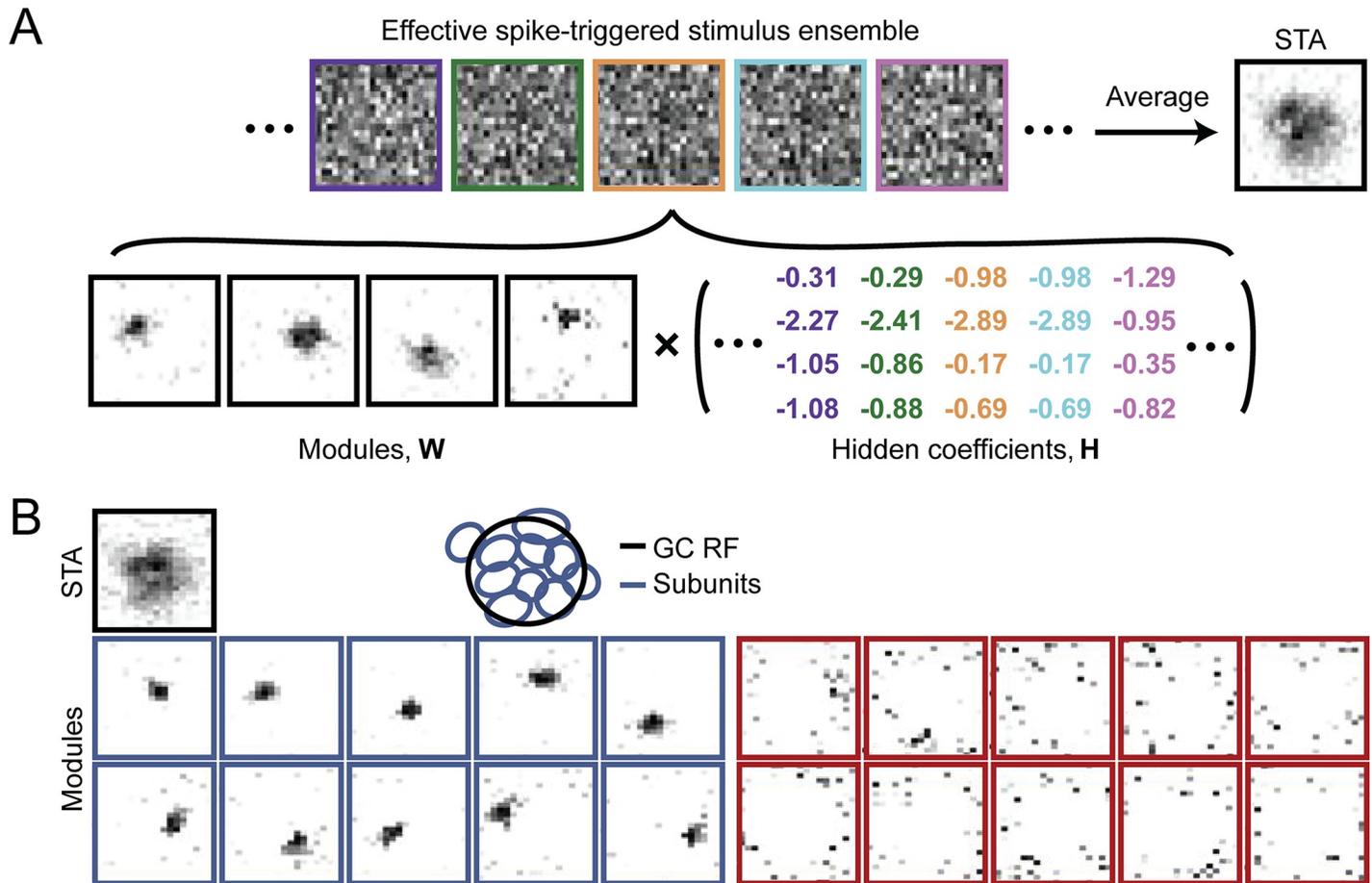


Fig 5. Identification of retinal ganglion cell subunits with STNMF. (A) Samples of a ganglion cell's effective spike-triggered stimulus ensemble (top), whose average corresponds to the cell's STA. For easier visual comparison with the subunits, STAs are displayed with negative pixel values set to zero and with zero corresponding to white in the grayscale image. STNMF decomposes this ensemble into a set of modules and hidden coefficients (bottom). The example here shows four modules that were identified for a sample ganglion cell. (B) Modules obtained for another sample ganglion cell by applying STNMF with 20 modules (bottom two rows). Some modules have a strongly localized structure (blue frames); others are more noise-like (red frames). These modules make up the subunits within a ganglion cell RF. The top row shows the cell's RF, given by the spatial component of the STA, as well as the fitted RF outline (GC RF, black ellipse), together with outlines of the localized subunits (blue ellipses). Scale bars, 100 μ m. GC, ganglion cell; RF, receptive field; STA, spike-triggered average; STNMF, spike-triggered nonnegative matrix factorization. Adapted with permission from [44].

<https://doi.org/10.1371/journal.pcbi.1006908.g005>

However, a number of recent studies suggest that responses are neither sparse nor low dimensional in V1 of the mouse [39, 40] and monkey [70]. Using high-density electrophysiology, Stringer and colleagues [40] found that the response of more than 10,000 visual cortical neurons to 2,000 image stimuli is high dimensional. In monkey V1, one needs to look at many principal components to decode natural images, and these principal components reflect contributions from most of the recorded neurons [70]. In addition, V1 neurons in the mouse might encode both visual stimuli and behavior in a mixed representation: a recent study found no separate sets of neurons encoding stimuli and behavioral variables, but each neuron multiplexed a unique combination of sensory and behavioral information [39]. These findings suggest that efficient coding might render an incomplete picture of sensory processing in V1 and that more studies are needed to reevaluate past findings. To this end, Stringer and colleagues [40] suggested that the population code of visual cortex might be determined by two constraints: efficiency, to make best use of the limited number of neurons, and smoothness, which allows similar stimuli to evoke similar responses.

In summary, there is a large body of research showing that computational models based on efficient coding, such as NSC, can account for a variety of response properties in early visual cortex. Although methods like spike-triggered average [71] and dimensionality reduction [72] give us confidence that we have a good understanding of the sensory response in V1, this understanding remains far from complete [73, 74] and in fact might be missing a number of dimensions related to task, state, or behavior [39, 40]. With the exception of face processing in IT [13, 19], NSC has yet to be applied to higher-order areas in the ventral visual pathway. The success of NSC in explaining V1 and V2 response properties suggests that it might be possible to extend the model to texture integration in V4.

NSC in the dorsal visual pathway

Our group found evidence for NSC in the dorsal subregion of the medial superior temporal (MSTd) area [46], which is part of the visual motion pathway in the dorsal visual stream. Neurons in MSTd respond to relatively large and complex patterns of retinal motion (“optic flow”), owing to input from direction- and speed-selective neurons in the middle temporal (MT) area (for a recent review, see [75]). Although MSTd had long been suspected to be involved in the analysis of self-motion, the complexity of neuronal response properties has made it difficult to experimentally investigate how neurons in MSTd might perform this function.

When our group applied NMF to simulated neural activity patterns whose statistical properties resembled that of experimentally recorded MT neurons [46], we found a sparse, parts-based representation of retinal flow (Fig 4B) similar to the parts-based representation of faces encountered by Lee and Seung [13]. The resulting “basis flow fields” showed a remarkable resemblance to RFs of MSTd neurons, as they responded to an intricate mixture of 3D translational and rotational flow components in a subset of the visual field. As a result, any flow field possibly to be encountered during self-movement through a 3D environment could be represented by only $B = 64$ simulated MSTd neurons, as compared with $F = 9,000$ simulated MT input neurons. This led to a sparse and parts-based population code in which any given stimulus could be represented by only a small number of simulated MSTd neurons [46].

Fig 6 shows the distribution of direction preferences of MSTd-like model units (Fig 6A and 6B; [46]) for rotation and translation, respectively. Each data point in the scatter plots specifies the preferred 3D direction of a model unit. Histograms along the boundaries show the marginal distributions of azimuth and elevation preferences. Not only did individual units match response properties of individual neurons in macaque MSTd [76]), but the model was able to recover statistical properties of the MSTd population as a whole, such as a relative overrepresentation of lateral headings.

MSTd is known to encode a number of perceptual variables, such as the direction of travel (heading) and eye rotation velocity. During forward movement, retinal flow radiates out symmetrically from a single point, the focus of expansion (FOE), from which heading can be inferred. However, instead of consisting of a set of distinct subpopulations, each specialized to encode a particular perceptual variable, MSTd has been found to consist of neurons that act more like basis functions, in which a majority of cells were involved in the simultaneous encoding of multiple perceptual variables (Fig 6C). A similar picture emerged when we investigated the involvement of MSTd-like model units in the encoding of both heading and eye rotation velocity (Fig 6C).

Interestingly, the sparsity regime in which model MSTd achieved the lowest heading prediction error (Fig 6D) was also the regime in which MSTd-like model units reproduced a variety of known MSTd visual response properties (for experimental details, refer to [46]). In

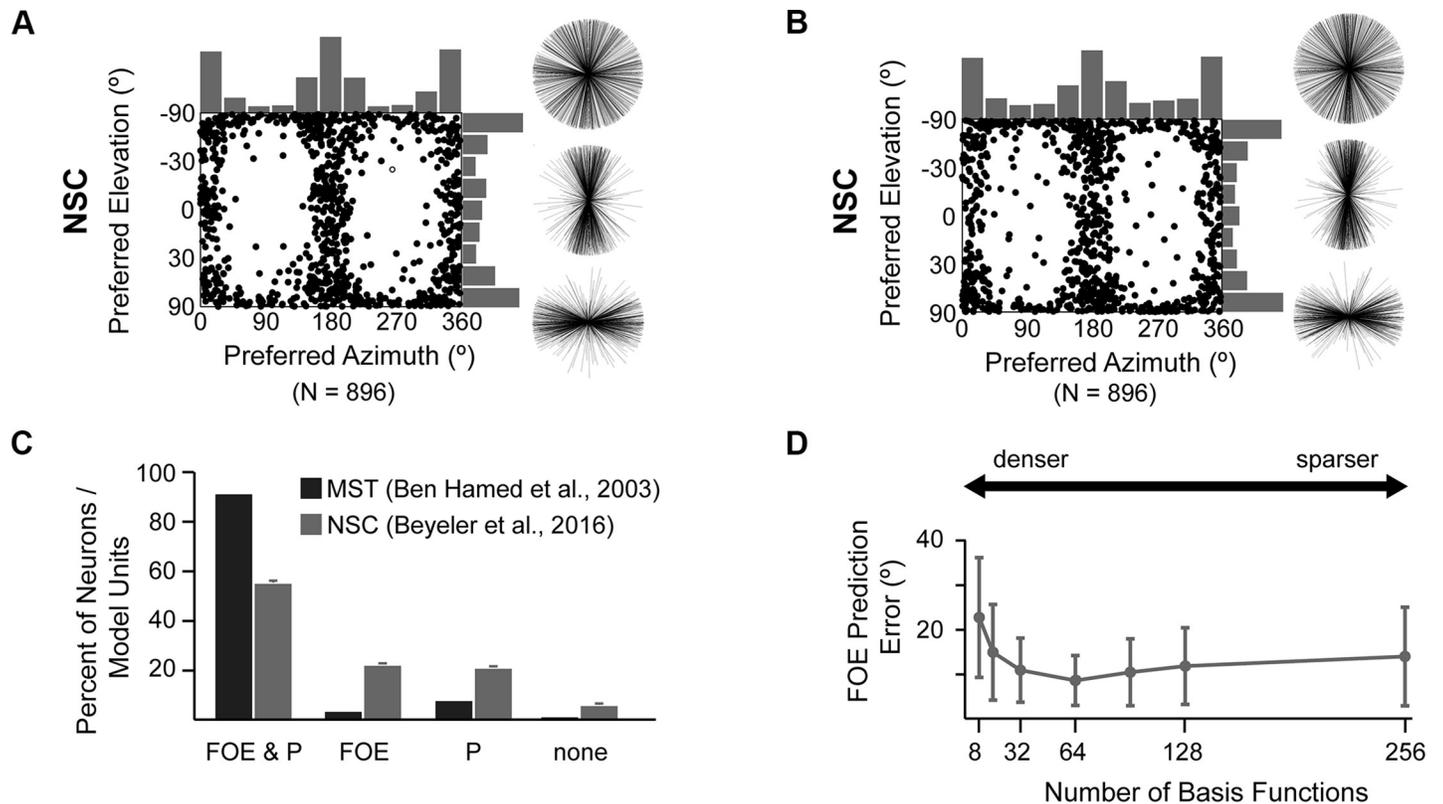


Fig 6. (A and B) Distribution of 3D direction preferences of MSTd-like model units in the NSC-based sparse decomposition model (rotation, [A]; translation, [B]). Each data point in the scatter plots corresponds to the preferred azimuth (abscissa) and elevation (ordinate) of a single neuron. Histograms along the top and right sides of each scatter plot show the marginal distributions. Also shown are 2D projections (front view, side view, and top view) of unit-length 3D preferred direction vectors (each radial line represents one neuron). (C) Distribution of FOE & P selectivities in macaque MSTd (dark gray) and model MSTd (light gray). Neurons or model units were involved in encoding heading (FOE), eye velocity (P), both (FOE & P), or neither (none). (D) Heading prediction (generalization) error as a function of the number of basis functions using cross validation. Vertical bars are the SD. FOE, focus of expansion; MSTd, dorsal subregion of the medial superior temporal area; NSC, nonnegative sparse coding; P, pursuit. *Reprinted with permission from [46].*

<https://doi.org/10.1371/journal.pcbi.1006908.g006>

contrast to findings about early visual cortex, this regime does not use an overcomplete basis set [35], yet it can still be considered a sparse coding regime [8] because only a few MSTd-like model units were needed to recover the stimulus, and each model unit responded to a subset of stimuli (see Fig 8C in [46]). Such an intermediary sparse code might be better suited (as opposed to an overcomplete basis set) for areas such as MSTd because the increased memory capacity of such a code might lead to compact and multifaceted encodings of various perceptual variables.

Taken together, the computational modeling work on MSTd described previously suggests that NSC is not specific to primary sensory areas and may be observed in other downstream sensory regions.

NSC in the auditory cortex

Analogous to early visual cortex, the auditory system is believed to decompose auditory signals into a set of elementary acoustic features [77] such that the complete acoustic waveform can be described by a sparse population code that operates near an information-theoretic optimum [77–79]. It is therefore not surprising that computational models based on NSC have been very successful at describing the spectro-temporal RF of neurons in the primary auditory cortex (A1) [80, 81]. Response properties of A1 neurons are well described by a spectrogram; they

are often tuned to stimulus frequency but are rarely phase locked to oscillations of the sound waveform [82]. The cortical representation of auditory signals seems to not only be sparse but also rely on statistically independent acoustic features [83].

Similar to visual cortex, auditory cortex is hierarchically organized, with neurons in A1 responding to simple acoustic features of natural sounds and higher-order areas responding to more behaviorally relevant stimuli. The anterior superior temporal region of auditory cortex, for example, responds to categories of acoustic objects, such as sounds produced by voices and musical instruments [82]. An intriguing question for future modeling studies is therefore whether NSC can be extended to the next level of the auditory hierarchy: Would it be possible to construct more complex acoustic objects from a sparse, parts-based set of elementary, A1-like acoustic features? And would the representation of such acoustic objects resemble neuronal responses in the anterior superior temporal region of auditory cortex?

Taken together, we suggest that auditory cortex is a good example for efficient and NSC-based coding in a sensory system other than the visual cortex, in which further study is warranted.

NSC in the olfactory cortex

The olfactory cortex is another nonvisual cortical area worth investigating for NSC-like responses. In contrast to most other sensory modalities, the basic perceptual dimensions of olfaction remain unclear. In particular, the olfactory modality is intrinsically high dimensional and lacks a simple, externally defined basis analogous to wavelength or pitch on which elemental odor stimuli can be quantitatively compared (for a recent review, see [84]). Odors evoke complex responses in granule cells (located in the olfactory bulb) that evolve over hundreds of milliseconds [85]. Granule cells use a sparse combinatorial code to convey information about odor identity and concentration [86, 87]. Downstream from the olfactory bulb, odors tend to activate a small but consistent proportion (approximately 10%) of cortical neurons in the piriform cortex [88], which is thought to form odor object percepts [89, 90]. Although piriform cortex is not topographically organized, a spatial structure can be discerned when examining the projections of output neurons, which are highly segregated and functionally specific. Whereas the anterior piriform cortex is associated with the encoding of odor identity and odor structure, the posterior piriform cortex is involved in associational aspects of odors, such as valence and similarity [89, 91].

A compelling piece of evidence for NSC in the olfactory system was recently provided by Castro and colleagues [48]: In an effort to elucidate the dimensions along which perceptual space might be organized in the olfactory system, they applied NMF to a perceptual dataset built from 144 monomolecular odors, each represented by a 146-dimensional vector (an “odor profile”). Each dimension in the odor profile corresponded to the rated applicability of a number of semantic labels, such as “sweet,” “floral,” and “heavy.” By applying NMF to the odor profile, they showed that a set of 10 sparsely activated basis functions could accurately describe any odor in the dataset (Fig 7A). Interestingly, NMF revealed a prominent block diagonal structure to the full matrix \mathbf{H} (Fig 7B), indicating that (1) a given odor tended to be characterized by a single prominent basis function, implying that the basis functions recovered by NMF were perceptually meaningful, and (2) all ten basis functions were being used approximately with equal frequency, implying that the basis functions recovered by NMF could span the space of behaviorally relevant odors. This suggests that a given odor percept may be considered an instance of one of several fundamental qualities.

Furthermore, NMF recovered basis functions whose descriptors aligned with perceptual dimensions highlighted in several previous analyses of odor space, including but not limited to

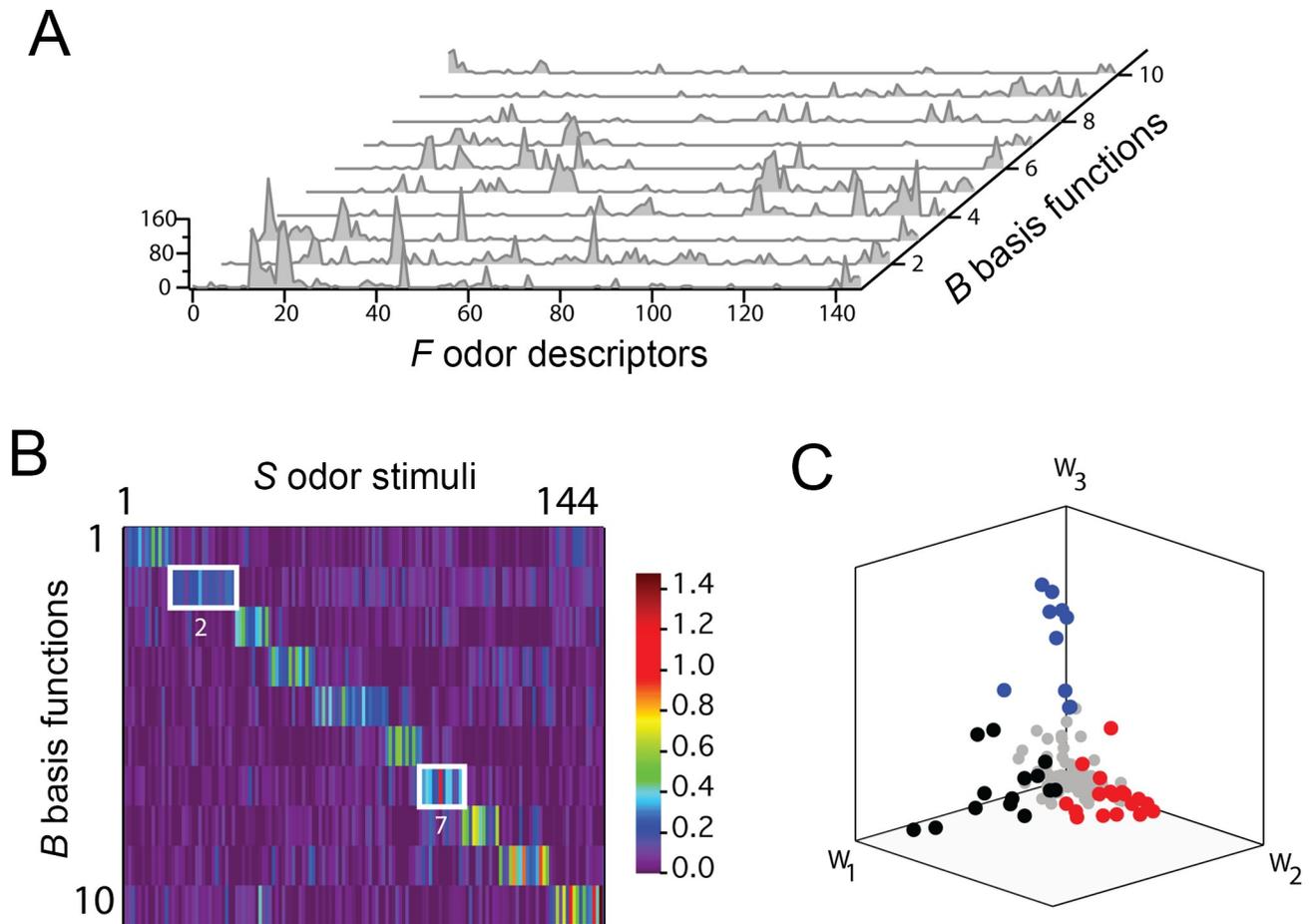


Fig 7. NMF recovers a sparse and parts-based representation of olfactory perceptual space. (A) Waterfall plot of the 10 basis functions constituting W (same nomenclature as in Fig 4). (B) Heat map of the hidden coefficient matrix, H , in which each column of H corresponds to a different odor. Columns of H are normalized and sorted. (C) Plot of all 144 odors in the dataset (each point is a column in H) in the space spanned by the first three basis functions, \vec{w}_1 (“fragrant”/“floral”), \vec{w}_2 (“woody, resinous”/“musty, earthy”), and \vec{w}_3 (“fruity, other than citrus”/“sweet”). Black, red, and blue points are those with their largest hidden coefficient corresponding to the first, second, and third basis function, respectively. Gray points are all remaining odors. Adapted with permission from [48].

<https://doi.org/10.1371/journal.pcbi.1006908.g007>

relative pleasantness (e.g., “fragrant,” “sickening”) and potential palatability (“woody, resinous,” “chemical,” “sweet,” and “lemon”). Odors clustered predominantly along these axes (as illustrated in Fig 7C) for three specific basis functions [48].

In summary, although sensory processing in the olfactory system remains an area of active research, there is evidence consistent with a sparse and parts-based encoding of odor identity and concentration. Only recently have NSC-based methods been employed to elucidate the neural code for olfaction. Future studies may provide additional supporting evidence.

NSC in the somatosensory cortex

In early areas of primary somatosensory cortex (S1), a number of parallels can be drawn to sparse, reduced information processing observed in other primary sensory cortices. First, activity in rodent barrel cortex, a region of S1 that is a major target for somatosensory inputs from the whiskers via the thalamus, can be extremely sparse [92–94], similar to activity in A1. Consequently, sparse-coding models have successfully explained the response properties of individual neurons in rat barrel cortex (e.g., Hafner and colleagues [95]). Second, similar to

V1, neurons in primate areas 3b and 1 of S1 act like Gabor filters for tactile orientation [96, 97]. The same is true for rat barrel cortex [98]. Third, similar to visual area MT, primate S1 contains a subpopulation of neurons that can infer the direction of tactile motion from a spatiotemporal pattern of activation across a 2D sensory sheet (i.e., the skin) [99]. Specifically, neurons in area 1 of S1 tend to respond to plaid textures in the same fashion that MT neurons respond to visual plaids [99]. These findings suggest that much of what can be said about sparse and parts-based information processing in visual cortex also applies to S1.

One NSC-like model that has enjoyed success in explaining complex S1 rodent response properties is the rectified latent variable model (RLVM), a combination of nonlinear dimensionality reduction with nonnegativity constraints. In an effort to elucidate the stimulus dimensions that individual S1 neurons respond to, Whiteway and Butts [100] applied RLVM to a two-photon imaging dataset of hundreds of simultaneously recorded neurons in mouse barrel cortex while the animal was performing a tactile discrimination task. Interestingly, they found basis functions that properly identified individual neurons. Similar to the recorded neuronal responses, these basis functions were closely related to both the tactile stimulation as well as nonstimulus aspects of the behavioral task. Furthermore, RLVM achieved a lower reconstruction error than other linear dimensionality reduction techniques such as **principal component analysis (PCA)**, thus highlighting the benefit of using NMF-based decompositions over PCA to explain neural data.

However, NSC has not been observed in nonhuman primate somatosensory cortex. Tactile information from various submodalities converges at later stages of monkey S1 [101, 102] and is multiplexed across different time scales using both rate and spike timing codes [103]. These regions might represent different stages in the processing pipeline leading to form and texture perception [104]. Primate area 2 of S1 is known to integrate both tactile and proprioceptive stimuli; for example, some neurons respond only to active reaching movements, some respond only to passive movements (e.g., unexpected perturbations to the hand that generate passive limb displacements), and others respond to both [105]. These complex response properties may argue against a sparse and parts-based code in area 2.

Taken together, neurons in early somatosensory cortex respond to a small number of stimulus dimensions, not unlike to sensory neurons in early visual and auditory cortex. However, current evidence argues against NSC in higher areas of somatosensory cortex. The parallels to the visual system are striking though: area 1, which resembles visual area MT by showing Gabor-like responses to tactile motion, feeds into area 2, which resembles visual area MSTd by showing intermingling of responses to tactile and proprioceptive stimuli (analogous to intermingling of visual and vestibular stimuli in MSTd). It is therefore not unthinkable that an NSC-like model that operates on neuronal inputs to area 2—constructed analogous to [46]—could reproduce some of these response properties. However, until the neuronal mechanisms underlying these complex response properties are better understood, one would have to conclude that NSC might not apply to later stages of somatosensory cortex.

NSC in the retrosplenial cortex

In our own work, we found evidence that NSC can explain response properties in RSC, an area important for navigation and spatial memory [106–108]. Neurons in the RSC conjunctively encode multiple variables related to the environment and one's position and movement within it (e.g., position, head direction, linear velocity, and angular velocity), allowing the representation of spatial features of the environment with respect to multiple reference frames [109].

Using a similar methodology to [46], we applied NMF, with a sparsity constraint, to parameterized behavioral variables extracted from electrophysiological recordings of RSC neurons in the rat [109] while the animal ran back and forth on a W-shaped track (for experimental details, see Supporting information). We found a sparse and parts-based representation for behaviorally relevant variables such as the animal's position, head direction, and movement direction (Fig 4C). Interestingly, model RSC neurons encoded these variables with respect to multiple frames of reference (e.g., head direction: **allocentric reference frame**, linear velocity: **route-based reference frame**). The dimensionality of the stimulus space was drastically reduced from $F = 417$ input neurons to a set of $B = 30$ model RSC neurons.

The basis functions recovered by NMF were then used to generate simulated responses of model RSC neurons according to Eq 4, and the simulated responses were compared with neuronal responses from the electrophysiological recordings. Interestingly, simulated neuronal activity could be classified into three broad categories, with remarkably similar population statistics to rat RSC: (1) responding to left and right turns on a specific position along the route, (2) responding to left and right turns regardless of the position along the route, and (3) exhibiting complex and robust firing patterns without turn sensitivity (see Fig 8A and 8B as well as Supporting information).

Taken together, this study suggests that neuronal population activity in RSC is consistent with NSC. This is an example that NSC can apply outside sensory cortex, even where responses have not traditionally been considered sparse or parts based.

Reinforcement-driven NSC in the basal ganglia

There is computational evidence for a reward-driven variant of NSC in the basal ganglia, a cluster of deep forebrain nuclei that are involved in the processing of motor, associative, and limbic information (for recent reviews, see [3, 110]). The basal ganglia network may be viewed as multiple parallel loops where cortical and subcortical projections interact with internal reentrant loops, forming a complex network ideally designed for selecting and inhibiting simultaneously occurring events and signals (for a recent review, see [111]). To achieve this function, the basal ganglia connect most cortical areas to the frontal cortex through a series of convergent and sparsely connected pathways [112], in which signals from tens of millions of cortical neurons are projected onto a 10–10,000-fold smaller population of neurons in different subnuclei of the basal ganglia [3]. Similar to the convergence of 100 million photoreceptors onto 1 million optic nerve fibers in the retina, these highly convergent pathways from cortex to the basal ganglia suggest a potential role for dimensionality reduction.

One possible model, termed the reinforcement-driven dimensionality reduction (RDDR) model, suggests that dimensionality reduction in the cortico-basal ganglia pathway is achieved via a combination of Hebbian and anti-Hebbian learning rules that are implemented by feed-forward excitatory and lateral inhibitory connections [3, 49]. These learning rules control the strength of synaptic weights in the network by altering the weight of a given synapse in proportion to the correlation between the firing rates of its presynaptic and postsynaptic neurons. In Hebbian learning, synaptic weights are strengthened given a positive correlation (leading to a phenomenon referred to as long-term potentiation [LTP]), whereas synaptic weights are depressed if the firing rate correlation is negative (leading to long-term depression [LTD]). On the other hand, in anti-Hebbian learning, which is typically applied to inhibitory connections, correlated activities are subjected to LTD, and uncorrelated activities are subjected to LTP. In order to implement dopamine-modulated Hebbian learning in this model, a reinforcement signal was used to dictate the level of dopamine in the circuit (1 for reward-related events, 0 for the absence of reward-related events, and negative values to simulate dopamine depletion)

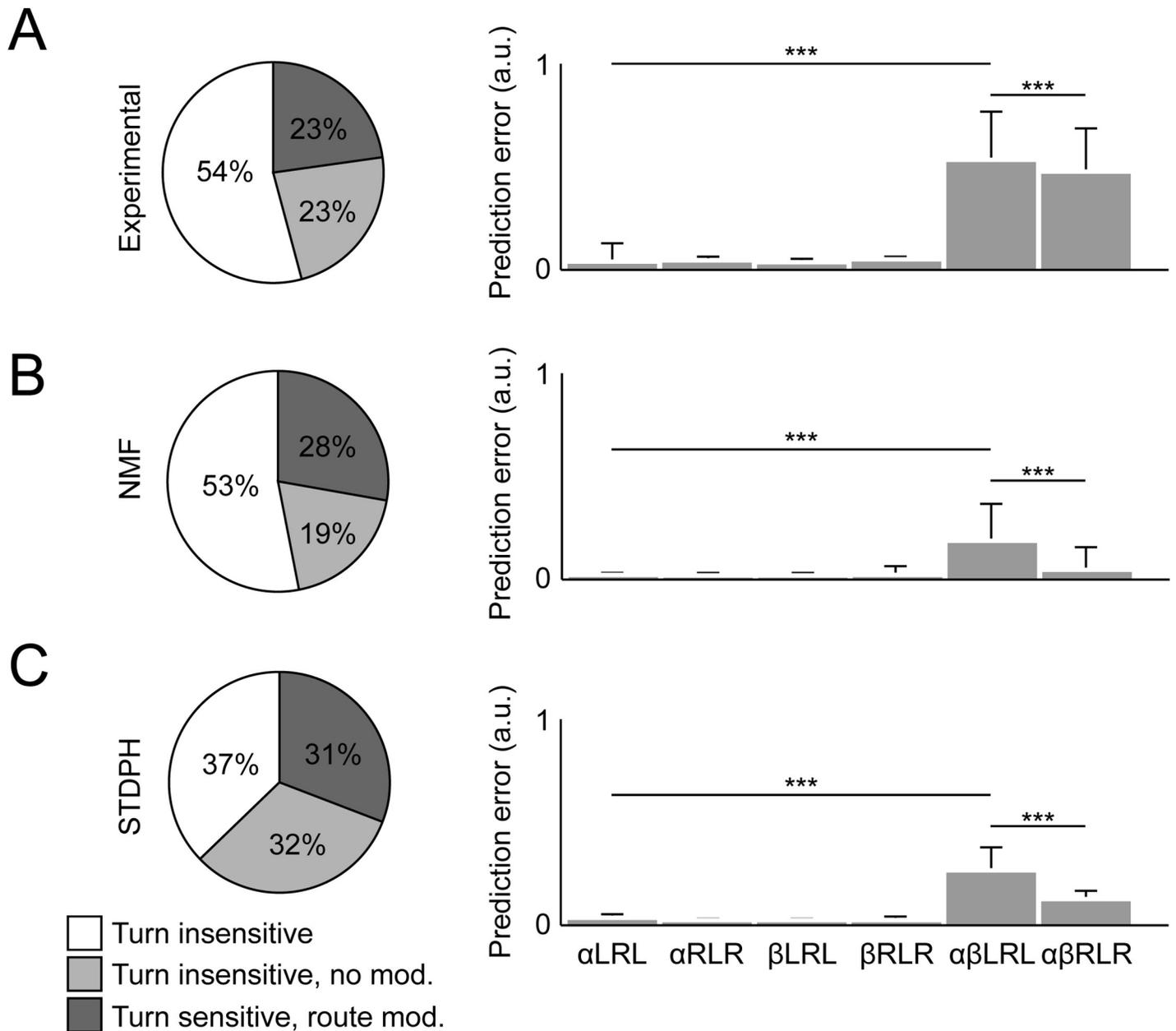


Fig 8. Comparison between experimental data and two computational models of rat RSC suggest a functional similarity between STDPH and NMF. Rats used two turn sequences (inbound: LRL; outbound: RLR) to traverse a W-shaped track located at two different allocentric locations (α , β). (A) Experimental data from [109]. (B) Simulated using NMF with sparsity constraints. (C) Simulated by evolving STDPH parameters to fit experimental data [127, 128]. Left column: Functional neuron type distributions. Right column: Location prediction errors. The prediction error is based on how well the neuronal population response can predict the rat's location on the maze. For details, see [50, 109]. Prediction error when comparing even and odd trials on the same maze in the same location in the room (prefix α or β) was much smaller than when the same maze was in different locations (prefix $\alpha\beta$; Kruskal-Wallis and Tukey's range test, *** = $p < 0.001$), demonstrating that the network can distinguish similar routes that occur in different allocentric positions. For details see Supporting information. LRL, left-right-left; NMF, nonnegative matrix factorization; RLR, right-left-right; RSC, retrosplenial cortex; STDPH, spike-timing-dependent plasticity and homeostatic synaptic scaling.

<https://doi.org/10.1371/journal.pcbi.1006908.g008>

[49]. The value of the reinforcement signal then determined the sign and magnitude of each synaptic weight change.

In the RDDR model, a reinforcement signal corresponding to dopamine modulates the Hebbian learning rule of the feedforward projections, allowing the network to learn to extract

input dimensions that are associated with reward activity while suppressing behaviorally irrelevant input dimensions. Whereas the original RDDR model was a neural network–based model for performing PCA [49], later iterations incorporated nonnegativity constraints on the connection weights that effectively transformed the model into an NMF variant [3]. The model predicted that these lateral connections facilitated learning by shaping correlations between neurons in the corticostriatal projections using dopamine-modulated LTP and LTD, which has yet to be experimentally validated.

In addition to suggesting a role for lateral connectivity in the basal ganglia, the RDDR model also advanced understanding of basal ganglia dysfunction in movement-related disorders such as Parkinson’s and Huntington’s disease. Previous studies had indicated that lesions to functionally healthy basal ganglia had minimal impact on behavior. Bar-Gad and colleagues [49] then suggested that this was an expected finding because of the network’s ability to reorganize connections, whereas abnormal dopamine levels should significantly alter the reinforcement signal that controls the model’s ability to discriminate behaviorally relevant input signals (as in Parkinson disease). Accordingly, restoration of background dopamine levels via dopamine replacement therapy alleviates the symptoms, consistent with results of dopamine depletion and restoration in the model.

In summary, NSC is a prime candidate to allow the basal ganglia to compress information in the cortico-basal ganglia pathway and extract input dimensions that are associated with reward activity. However, the complexity of the basal ganglia network has so far prohibited a deep scientific understanding of the multifaceted neural computations it performs.

Discussion

NSC in the brain

We reviewed compelling evidence that a wide range of neuronal responses can be understood as an emergent property of efficient coding due to dimensionality reduction and sparsity constraints. In particular, NSC might be employed by sensory areas to efficiently encode external stimulus spaces, by some associative areas to conjunctively represent multiple behaviorally relevant variables, and possibly by the basal ganglia to coordinate movement.

NSC is tightly connected to a number of unsupervised learning techniques, such as NMF (a popular tool for high-dimensional data analysis [113]), *k*-means clustering (an algorithm used to partition *n* observations into *k* clusters [114]), and **independent component analysis (ICA)** (a computational method for separating a multivariate signal into additive, statistically independent subcomponents). Both NMF and ICA are capable of decomposing high-dimensional data into parts-based representations—in contrast to PCA, which usually results in holistic representations [13]. As originally noted by Hoyer [42], if the fixed-norm constraint is placed on the rows of **H** instead of the columns of **W**, Eq 3 can be directly interpreted as the joint log-posterior of the basis functions and hidden components in the noisy ICA model [64].

Similarly, NSC is closely related to compressed sensing (for a recent review, see [115]), and a recent study has even suggested to combine the two [116]. Compressed sensing posits that neurons might implement dimensionality reduction by randomly projecting patterns of activity into a lower-dimensional space—namely, by synaptically mapping *N* upstream neurons to a downstream region containing *M* < *N* neurons. Analogously, compressed sensing supports dimensionality expansion by projecting into a larger downstream area [115]. The theory of compressed sensing then provides the mathematical tools to reconstruct the original space from the random projections.

NSC, ICA, and compressed sensing often make similar predictions that only slightly differ in the nature of the basis function representation necessary to achieve optimal reconstruction

(for details, please refer to the Discussion of [115]). For example, whereas ICA emphasizes the statistical independence of unmixed sources, and compressed sensing requires basis function to be “maximally incoherent” [115], NSC does not make any such assumptions as long as the basis functions are nonnegative.

Potential neural mechanisms. To operate efficiently, it has been suggested that the brain might enforce geometrical and biophysical constraints on axonal wiring [117]. In addition to reducing overall wiring length [118], the brain might also aim to minimize local delays by favoring a high degree of local connectivity [119]. If connectivity reflects coding [120], it would not be surprising to find that such ecological considerations carry over into brain function, favoring sparse population codes and neuronal representations that are local in functional space (i.e., parts-based). However, wiring cost is likely to be only one of many constraints on the brain connectome, perhaps supplementing competitive pressures for hub-mediated information exchange between network modules [121].

In addition, evidence suggests that Hebbian-like synaptic plasticity rules allow neurons to perform statistical inference on their inputs [47, 122–124]. One particular study demonstrated through a mathematical proof that a certain form of **spike-timing-dependent plasticity (STDP)** in combination with homeostatic synaptic scaling (i.e., STDP and homeostatic synaptic scaling [STDPH]) can approximate the NMF algorithm [123]. Similar to Oja’s rule [124], which was developed to stabilize rate-based Hebbian learning (effectively resulting in PCA), Carlson and colleagues showed that synaptic scaling acts as a homeostatic mechanism to stabilize STDP (effectively resulting in NMF). Interestingly, we were able to apply these ideas to electrophysiologically recorded neuronal activity observed in the RSC of rats during a spatial navigation task (Fig 8; for experimental details, see Supporting information). Both STDPH and NMF were able to recover key features such as encoding spatial reference frames (i.e., allocentric and route-centric firing patterns) and position discrimination by reducing the dimensionality of behavioral variables (e.g., velocity, head direction, position). The neuronal and population responses from NMF and STDPH were comparable to the experimental findings [109]. Furthermore, the STDPH model contained a highly flexible and generalizable code that could automatically encode new routes through the same environment without retraining [50].

However, more research is needed to elucidate any potential connection between NSC and the many different synaptic plasticity rules commonly found across brain regions, different stages in the life of an animal, and animal species (e.g., [125, 126]).

Model limitations. Although NSC has proved useful in understanding a variety of neuronal responses as an emergent property of efficient population coding based on dimensionality reduction and sparse coding, it is clear that it does not apply to everywhere in the brain. In this section, we discuss some of the limitations to this theory.

First, there is increasing evidence that several motor variables are encoded throughout the brain, including early sensory areas [39, 129–131]. For example, in the mouse, running modulates the gain of visual inputs [132, 133] and is critical for integration of visual motion [134, 135]. A potential explanation for these widespread effects is that certain movements reflect changes in the animal’s internal state, such as increased arousal during running [133]. Interestingly, these results were not specific to visual cortex but were seen across wide regions of the mouse forebrain, suggesting that neuronal activity in these areas is more than just an efficient encoding of sensory stimuli.

Second, a practical limitation of dimensionality analyses in general is that the apparent dimensionality of the population response changes systematically with the complexity of the input space [8, 72, 136]. For practical purposes, simulated models of neuronal circuitry are typically built with far fewer units than the number of neurons in the real network. By excluding

input dimensions that are present in the brain, one will implicitly guide the simulated model away from spurious interactions that the real circuitry would have to handle. As a result, the simulated model might underestimate the true complexity of the task. Rather than trying to put an absolute number to the dimensionality of neuronal population activity in a given brain area, one should instead systematically vary the input to said brain area and ask whether the outputs of dimensionality reduction change in a sensible manner [72].

Third, part of the confusion about the role of sparse coding in the brain may arise from the wide variety of definitions of sparsity used in the literature [8, 38]: sparse population activity does not necessarily imply a sparse-coding scheme. In its widest possible theoretical sense, a neuron population exhibits sparse activity if the average activation ratio remains below 50% for binary neurons or below 100% for thresholded, rate-based neurons [8]. However, it is not surprising that different brain areas might employ different degrees of sparsity. For example, a network of “grandmother” cells might be able to implement a maximally sparse code, but it would have to do so by sacrificing representational capacity (as in Fig 1) and fault tolerance (i.e., the capacity to handle neuronal noise or the loss of a subset of neurons) [5, 8]. Instead, some brain areas might prefer to operate at a degree of sparsity that still allows for some level of robustness or adaptability. It is conceivable that this “point of operation” might depend on the complexity of the stimulus to be encoded or the task to be performed—for example, favoring an extremely sparse code in V1 [35] but giving rise to a slightly denser code with greater representational capacity in higher-order visual areas such as MSTd, which could lead to compact and multifaceted encodings of various perceptual variables (see Discussion in [46]).

Furthermore, NSC does not apply to many regions of the brain, especially when the role of the brain area is to cause behavior. Two prominent examples in which NSC has not been observed are the prefrontal cortex (PFC) and motor cortex. For one, studies indicate that the population code in these regions may be quite dense (as in Fig 1A). For another, dimensionality reduction studies in these regions suggest that individually neurons typically encode multiple task-related signals at once, such as the animal’s upcoming choice, state, and the strength of sensory evidence [14, 15, 137–140]. Whereas feedforward neural networks are better predictors of neuronal processing in early sensory areas, it is interesting to note that motor activity seems to be well represented by recurrent neural networks [141] as well as models based on random projection theory [142].

There is increasing evidence that responses in the primary motor cortex (M1) are neither sparse nor parts based [137, 138, 143–146]. Instead, M1 neurons tend to be active during most movements [145], and response patterns of individual neurons rarely match those of individual muscles [138, 144]. In addition, M1 neurons show temporally complex patterns of activity leading to high-dimensional population activity [138], as measured by PCA during a reaching task. These findings might be due to the role of the motor system, which is to cause behavior rather than represent features. Motor cortex can be understood as part of a larger dynamical system spanning many areas, including the spinal cord, and incorporating sensory feedback (for a recent review, see [146]). In this view, correlations between activity and movement parameters need not represent anything so long as the right patterns of activity are created at the level of the spinal cord [147–149].

Potential for NSC in other brain regions

There are several nonsensory areas that may demonstrate NSC. In this section we point to evidence that suggests this is the case but also discuss how sparse activity in these regions differs from NSC in sensory systems. We suggest that further studies should be carried out to assess the potential for NSC in these regions.

Hippocampus. There are aspects of the hippocampus that are comparable to the ideas of NSC proposed here. For example, the dentate gyrus is often associated with sparse activity [150–152]. The expansion from a dense, entorhinal cortex coding to sparse dentate gyrus activity suggested a mechanism of pattern separation [153, 154]. Since these early theories, accumulating evidence has supported the role of the dentate gyrus in pattern separation [150, 152, 155]. However, the dentate gyrus projects to a highly recurrent CA3 region, which does not appear to be sparse or reduce dimensionality. Rather, the CA3 region has a role in pattern completion through autoassociation [153, 154, 156]. Another feature of hippocampal processing that is related to NSC may be the place cell itself. Sparsity has long been used as a metric for place cell quality [157]. A place cell is said to be sparse if it fires in a small region of the environment, and this is related to how much spatial information that cell encodes. In a sense, each place cell can be thought of as a "part" of the environment, and these parts cover the entire environment. Although this evidence points to sparse, parts-based, and reduced coding in some hippocampal regions, it does differ from the NSC representations in sensory areas. That is, sparse coding as defined previously for sensory areas is not necessarily the same as sparsity or as sparse activity leading to pattern separation. Still, it would be of interest to apply NSC to the hippocampal inputs, similar to the methods applied to RSC during a spatial navigation task [50], to see if sparse, reduced basis functions emerge in other regions.

Parietal cortex. Analogous to our modeling work in MSTd [46], it might be possible to apply NSC to other areas of the posterior parietal cortex that are involved in multisensory heading perception. Areas such as the ventral intraparietal (VIP) area and the visual posterior sylvian (VPS) area are also known to respond to optic flow, but they increasingly respond to inertial vestibular stimulation as well [158]. Although the degree of sparsity of the population code in VIP and VPS is not known, the fact that neurons in these areas respond to mixtures of visual and vestibular heading cues make them prime examples to be examined with an NSC-based modeling approach.

Elsewhere in parietal cortex, single neurons act as basis functions to represent the spatial configuration of objects with respect to multiple reference frames (e.g., by transforming eye-centered to body-centered coordinates) [18, 159, 160]. This is similar to the integration of multimodal heading cues mentioned previously, as well as to other associative areas such as RSC, which demonstrates conjunctive coding of various spatial navigation cues [50, 109]. There is further evidence that actions are represented in parietal cortex with respect to arbitrary and abstract reference frames, such as with respect to a planned route through an environment [161]. From a theoretical standpoint, NSC seems a good candidate to find an efficient, reference frame-agnostic representation of various behaviorally relevant variables [27, 45, 162, 163], but future studies will have to address these issues step by step.

Future directions

In addition to the areas highlighted previously, the essential components of NSC might be present in other brain regions not traditionally associated with the efficient encoding of information. We offer three testable predictions of this theory:

First, we suggest that a variety of neuronal response properties can be understood as an emergent property of efficient population coding based on dimensionality reduction. Depending on input stimulus and task complexity, we expect the dimensionality of the population code to be adjusted according to the bias–variance dilemma (Fig 3). This point of operation might differ across brain areas—for example, favoring neurons that respond to a small number of stimulus dimensions in V1 [35] but giving rise to mixed selectivity in higher-order brain areas such as MSTd [46] and RSC [50, 164].

Second, we predict that parts-based representations can explain RFs of neurons in a variety of sensory and associative cortices, including but not limited to those brain areas discussed here. In agreement with the literature on basis function representations [18, 159, 160], we expect parts-based representations to be prevalent in regions where neurons exhibit a range of tuning behaviors [46], display mixed selectivity [165, 166], or encode information in multiple reference frames [50, 109, 164].

Third, where such representations occur, we expect the resulting neuronal population activity to be relatively sparse in order to encode information both accurately and efficiently. As noted previously, sparse codes offer a trade-off between dense codes (in which every neuron is involved in every context, leading to great memory capacity but suffering from cross talk among neurons) and local codes (in which there is no interference but also no capacity for generalization).

Conclusion

In conclusion, there is increasing evidence that NSC can account for neuronal response properties in a number of sensory and associative cortices, as well as subcortical areas such as the basal ganglia. Although NSC might not apply to all brain areas—for example, motor or executive function areas—the success of NSC-based models, especially in sensory areas, warrants further investigation for neural correlates in other regions.

Data availability

The software used to generate some of the data presented in Figs 1B, 2 and 6A is archived on Zenodo ([10.5281/zenodo.2641351](https://doi.org/10.5281/zenodo.2641351)). The latest version is available on GitHub: <https://github.com/mbeyeler/2019-nonnegative-sparse-coding>.

Glossary

- **Allocentric reference frame.** A spatial frame of reference that is defined with respect to a broader environment (e.g., one's location on a map). Hippocampal place cells are a textbook example of neurons that are anchored to an allocentric reference frame
- **Basis functions.** A lower-dimensional set of linearly independent elements that can represent a high-dimensional input space given a weighted sum of these elements, in which the weight of each element is defined by a separate hidden component. For example, according to Fourier analysis, sine and cosine are basis functions for the space of all continuous periodic functions.
- **Dimensionality reduction.** The process of reducing the dimensionality of a space to the lowest possible space that encapsulates the variance of the original data via feature extraction. In the case of neuronal firing rate patterns, this means representing all possible firing rate patterns in the brain region using the smallest possible subset of the neurons
- **Efficient coding.** A theoretical model of sensory coding in the brain based on information theory [24–26]. The efficient coding hypothesis posits that sensory pathways can be understood as communication channels in which neuronal spiking aims to maximize available channel capacity by minimizing the redundancy between representational units
- **Holistic representation.** Representation of a stimulus space that does not rely on explicit representations of stimulus component parts. For example, a house might be represented by the visual system as a set of house “templates.” Although visual information from individual

house components (e.g., front door, windows, roof, etc.) would of course be included in the house representation, that information would not be contained in representational packets corresponding to the parse of the house into these features. Instead, houses would be recognized “all of a piece.”

- **ICA.** A computational method for decomposing multivariate data into additive components by assuming that the components are non-Gaussian signals and statistically independent from each other. Independent components differ from decorrelated components by the fact that the minimization includes higher-order and not only second-order statistics. A simple application of ICA is the “cocktail party problem,” in which the underlying speech signals are separated from a sample data consisting of people talking simultaneously in a room
- **NMF.** A computational method for decomposing multivariate data into additive components by constraining the components to be nonnegative. This constraint results in a parts-based representation because it only allows additive and not subtractive combinations of subcomponents
- **Parts-based representation.** Representation of a stimulus space in terms of explicit representations of stimulus component parts. For example, a house might be decomposed by the visual system into a set of doors, windows, a roof, etc. The resulting representation of the house would consist of representations of these parts, somehow linked together
- **PCA.** A computational method for decomposing multivariate data into linearly uncorrelated components. PCA identifies an ordered set of orthogonal directions that captures the greatest variance in the data [14]
- **Representational capacity.** The number of recognizably different patterns of neuronal activity that a population of neurons can generate in a useful time interval [167]
- **Route-centric reference frame.** A spatial frame of reference that is defined with respect to a planned path through a broader environment. For example, neurons in some parts of the brain fire for a particular location in a route, even if the route is repositioned or reoriented in the broader environment [161]
- **Sparse coding.** A population coding scheme in which activity is represented by the strong activation of a relatively small set of neurons. Sparse coding can be described as a trade-off between the benefits and drawbacks of dense and local codes [5]
- **STDP.** A Hebbian-inspired learning rule in which weight updates are computed based on the precise spike times of pre- and postsynaptic neurons that induce either LTP or LTD in the synapse, depending on whether the total presynaptic spike count preceded the total postsynaptic spike count, integrated over a critical temporal window

Supporting information

S1 Supplemental Information. Experimental details: Dimensionality reduction in RSC. RSC, retrosplenial cortex.
(PDF)

S1 Fig. SNN architecture used in the RSC experiment. The SNN had four input groups and a total of 417 excitatory input neurons (390 neurons for position, eight for head direction, 12 for linear velocity, and seven for angular velocity). There were a total of 600 output Izhikevich neurons, of which 80% (480) were excitatory neurons and 20% (120) were inhibitory neurons.

Network connectivity was set at 10% probability across all connections (inp → inh, inp → exc, inh → exc, and exc ↔ exc), and each connection type was governed by its own STDPH curve (excitatory STDP on the inp → exc, inh → exc, and exc ↔ exc connections and inhibitory STDP on the inp → inh connections). exc, excitatory; inh, inhibitory; inp: input; RSC, retrosplenial cortex; SNN, spiking neural network; STDPH, spike-timing-dependent plasticity with homeostatic synaptic scaling.

(PDF)

References

1. Koch C. Biophysics of computation: Information processing in single neurons. Cary, NC: Oxford University Press; 1999.
2. Kempermann G. Why New Neurons? Possible Functions for Adult Hippocampal Neurogenesis. *Journal of Neuroscience*. 2002; 22(3):635–638. PMID: [11826092](https://pubmed.ncbi.nlm.nih.gov/11826092/)
3. Bar-Gad I, Morris G, Bergman H. Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Progress in Neurobiology*. 2003; 71(6):439–473. <https://doi.org/10.1016/j.pneurobio.2003.12.001> PMID: [15013228](https://pubmed.ncbi.nlm.nih.gov/15013228/)
4. Babinsky R, Calabrese P, Durwen H, Markowitsch H, Brechtelsbauer D, Heuser L, et al. The possible contribution of the amygdala to memory. *Behavioural Neurology*. 1993; 6(3):167–170. <https://doi.org/10.3233/BEN-1993-6310> PMID: [24487116](https://pubmed.ncbi.nlm.nih.gov/24487116/)
5. Földiák P. Forming sparse representations by local anti-Hebbian learning. *Biological cybernetics*. 1990; 64(2):165–170. PMID: [2291903](https://pubmed.ncbi.nlm.nih.gov/2291903/)
6. Field DJ. What is the goal of sensory coding? *Neural computation*. 1994; 6:559–601.
7. Levy WB, Baxter RA. Energy-efficient neural codes. *Neural Computation*. 1996; 8:531–543. PMID: [8868566](https://pubmed.ncbi.nlm.nih.gov/8868566/)
8. Spanne A, Jorntell H. Questioning the role of sparse coding in the brain. *Trends in Neurosciences*. 2015; 38(7):417–427. <https://doi.org/10.1016/j.tins.2015.05.005> PMID: [26093844](https://pubmed.ncbi.nlm.nih.gov/26093844/)
9. Vinje WE, Gallant JL. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*. 2000; 287:1273–1276. PMID: [10678835](https://pubmed.ncbi.nlm.nih.gov/10678835/)
10. Rolls ET, Treves A. The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Network*. 1990; 1:407–421.
11. Tanaka JW, Farah MJ. Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*. 1993; 46A(2):225–245.
12. Palmer SE. Hierarchical structure in perceptual representation. *Cognitive Psychology*. 1977; 9:441–474.
13. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature*. 1999; 401(6755):788–91. <https://doi.org/10.1038/44565> PMID: [10548103](https://pubmed.ncbi.nlm.nih.gov/10548103/)
14. Cunningham JP, Yu BM. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*. 2014; 17:1500–1509. <https://doi.org/10.1038/nn.3776> PMID: [25151264](https://pubmed.ncbi.nlm.nih.gov/25151264/)
15. Rigotti M, Barak O, Warden MR, Wang XJ, Daw ND, Miller EK, et al. The importance of mixed selectivity in complex cognitive tasks. *Nature*. 2013; 497:585–590. <https://doi.org/10.1038/nature12160> PMID: [23685452](https://pubmed.ncbi.nlm.nih.gov/23685452/)
16. Park IM, Meister ML, Huk AC, Pillow JW. Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nature Neuroscience*. 2014; 17:1395–1403. <https://doi.org/10.1038/nn.3800> PMID: [25174005](https://pubmed.ncbi.nlm.nih.gov/25174005/)
17. Pagan M, Rust NC. Quantifying the signals contained in heterogeneous neural responses and determining their relationships with task performance. *Journal of Neurophysiology*. 2014; 112:1584–1598. <https://doi.org/10.1152/jn.00260.2014> PMID: [24920017](https://pubmed.ncbi.nlm.nih.gov/24920017/)
18. Pouget A, Sejnowski TJ. Spatial transformations in the parietal cortex using basis functions. *Journal of Cognitive Neuroscience*. 1997; 9(2):222–237. <https://doi.org/10.1162/jocn.1997.9.2.222> PMID: [23962013](https://pubmed.ncbi.nlm.nih.gov/23962013/)
19. Chang L, Tsao DY. The code for facial identity in the primate brain. *Cell*. 2017; 169:1013–1028.e14. <https://doi.org/10.1016/j.cell.2017.05.011> PMID: [28575666](https://pubmed.ncbi.nlm.nih.gov/28575666/)
20. Brunton BW, Johnson LA, Ojemann JG, Kutz JN. Extracting spatio-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of Neuroscience Methods*. 2016; 258:1–15. <https://doi.org/10.1016/j.jneumeth.2015.10.010> PMID: [26529367](https://pubmed.ncbi.nlm.nih.gov/26529367/)

21. Pillow JW, Simoncelli EP. Dimensionality reduction in neural models: An information-theoretic generalization of spike-triggered average and covariance analysis. *Journal of Vision*. 2006; 6(4):9. <https://doi.org/10.1167/6.4.9> PMID: 16889478
22. Sharpee T, Rust NC, Bialek W. Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Computation*. 2014; 16:223–250.
23. Gao P, Trautmann E, Yu BM, Santhanam G, Ryu S, Shenoy K, et al. A theory of multineuronal dimensionality, dynamics and measurement. *BioRxiv [Preprint]*. 2017. <https://doi.org/10.1101/214262>
24. Barlow HB. Possible principles underlying the transformation of sensory messages. In: Rosenblith WA, editor. *Sensory communication*. Cambridge, MA: MIT Press; 1961.
25. Attneave F. Some informational aspects of visual perception. *Psychology Review*. 1954; 61:183–193.
26. Linsker R. Perceptual neural organization: some approaches based on network models and information theory. *Annual Review of Neuroscience*. 1990; 13(1):257–281.
27. Louie K, Glimcher PW. Efficient coding and the neural representation of value. *Annals of the New York Academy of Sciences*. 2012; 1251:13–32. <https://doi.org/10.1111/j.1749-6632.2012.06496.x> PMID: 22694213
28. Simoncelli EP, Olshausen BA. Natural image statistics and neural representation. *Annual Review of Neuroscience*. 2001; 24:1193–1216. <https://doi.org/10.1146/annurev.neuro.24.1.1193> PMID: 11520932
29. Laughlin S. A simple coding procedure enhances neuron's information capacity. *Zeitschrift für Naturforschung C*. 1981; 36:910–912.
30. Pitkow X, Meister M. Decorrelation and efficient coding by retinal ganglion cells. *Nature Neuroscience*. 2012; 15:628–635. <https://doi.org/10.1038/nn.3064> PMID: 22406548
31. Dan Y, Atick JJ, Reid RC. Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory. *Journal of Neuroscience*. 1996; 16:3351–3362. PMID: 8627371
32. Shannon CE, Weaver W. *The mathematical theory of communication*. Champaign, IL: University of Illinois Press, Urbana; 1949.
33. Simoncelli EP. Vision and the statistics of the visual environment. *Current Opinion in Neurobiology*. 2003; 13:144–149. PMID: 12744966
34. Olshausen BA, Field DJ. Natural image statistics and efficient coding. *Network*. 1996; 7(2):333–339. <https://doi.org/10.1088/0954-898X/7/2/014> PMID: 16754394
35. Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*. 1996; 381:607–609. <https://doi.org/10.1038/381607a0> PMID: 8637596
36. Hyvärinen A, Hoyer PO. A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Research*. 2001; 41:2413–2423. PMID: 11459597
37. Olshausen BA, Field DJ. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision research*. 1997; 37(23):3311–3325. PMID: 9425546
38. Barth AL, Poulet JF. Experimental evidence for sparse firing in the neocortex. *Trends in Neuroscience*. 2012; 35:345–355.
39. Stringer C, Pachitariu M, Steinmetz N, Reddy C, Carandini M, Harris KD. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*. 2019; 364(6437):1–11. <https://doi.org/10.1126/science.aav7893> PMID: 31000656
40. Stringer C, Pachitariu M, Steinmetz N, Carandini M, Harris KD. High-dimensional geometry of population responses in visual cortex. *BioRxiv [Preprint]*. 2018. <https://doi.org/10.1101/374090>
41. Hoyer PO. Modeling receptive fields with non-negative sparse coding. *Neurocomputing*. 2003; 52–54:547–552.
42. Hoyer PO. Non-negative sparse coding. In: *Proceedings of the 2002 12th IEEE Workshop on Neural Networks for Signal Processing; 2002; Martigny, Switzerland*. Piscataway, NJ: IEEE; 2002. p. 557–565.
43. Beyeler M, Dutt N, Krichmar JL. Sparse and efficient neuromorphic population coding. *US Patent App 15/417626*. 2017.
44. Liu JK, Schreyer HM, Onken A, Rozenblit F, Khani MH, Krishnamoorthy V, et al. Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization. *Nature Communications*. 2017; 8(1):149. <https://doi.org/10.1038/s41467-017-00156-9> PMID: 28747662
45. Ben Hamed S, Page W, Duffy C, Pouget A. MSTd neuronal basis functions for the population encoding of heading direction. *Journal of Neurophysiology*. 2003; 90:549–558. <https://doi.org/10.1152/jn.00639.2002> PMID: 12750416

46. Beyeler M, Dutt N, Krichmar JL. 3D Visual Response Properties of MSTd Emerge from an Efficient, Sparse Population Code. *The Journal of Neuroscience*. 2016; 36(32):8399–8415. <https://doi.org/10.1523/JNEUROSCI.0396-16.2016> PMID: 27511012
47. Moreno-Bote R, Drugowitsch J. Causal inference and explaining away in a spiking network. *Scientific Reports*. 2015; 5:1–18. <https://doi.org/10.1038/srep17531> PMID: 26621426
48. Castro JB, Ramanathan A, Chennubhtola CS. Categorical dimensions of human odor descriptor space revealed by non-negative matrix factorization. *PLoS ONE*. 2013; 8:1–16.
49. Bar-Gad I, Havazelet-Heimer G, Goldberg JA, Ruppin E, Bergman H. Reinforcement-driven dimensionality reduction—a model for information processing in the basal ganglia. *J Basic Clin Physiol Pharmacol*. 2000; 11:305–320. PMID: 11248944
50. Rounds EL, Alexander AS, Nitz DA, Krichmar JL. Conjunctive Coding in an Evolved Spiking Model of Retrosplenial Cortex. *Behavioral Neuroscience*. 2018; 132(5):430–452. <https://doi.org/10.1037/bne0000236> PMID: 29863371
51. Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111:8619–8624. <https://doi.org/10.1073/pnas.1403112111> PMID: 24812127
52. Tanaka JW, Simonyi D. The "parts and wholes" of face recognition: a review of the literature. *The Quarterly Journal of Experimental Psychology*. 2016; 69(10):1876–1889. <https://doi.org/10.1080/17470218.2016.1146780> PMID: 26886495
53. Tsao DY, Freiwald WA, Tootell RBH, Livingstone MS. A cortical region consisting entirely of face-selective cells. *Science*. 2006; 311:670–674. <https://doi.org/10.1126/science.1119983> PMID: 16456083
54. Grimaldi P, Saleem KS, Tsao D. Anatomical connections of the functionally defined "face patches" in the macaque monkey. *Neuron*. 2016; 90:1325–1342. <https://doi.org/10.1016/j.neuron.2016.05.009> PMID: 27263973
55. Freiwald WA, Tsao DY, Livingstone MS. A face feature space in the macaque temporal lobe. *Nature Neuroscience*. 2009; 12:1187–1196. <https://doi.org/10.1038/nn.2363> PMID: 19668199
56. Tsunoda K, Yamane Y, Nishizaki M, Tanifuji M. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*. 2001; 4:832–838. <https://doi.org/10.1038/90547> PMID: 11477430
57. Issa EB, DiCarlo JJ. Precedence of the eye region in neural processing of faces. *Journal of Neuroscience*. 2012; 32:16666–16682. <https://doi.org/10.1523/JNEUROSCI.2391-12.2012> PMID: 23175821
58. Majaj NJ, Hong H, Solomon EA, DiCarlo JJ. Simple learned weighted sums of inferior temporal neuronal firing rates accurately predict human core object recognition performance. *Journal of Neuroscience*. 2015; 35:13402–13418. <https://doi.org/10.1523/JNEUROSCI.5181-14.2015> PMID: 26424887
59. Brincat SL, Connor CE. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience*. 2004; 7:880–886. <https://doi.org/10.1038/nn1278> PMID: 15235606
60. Popivanov ID, Schyns PG, Vogels R. Stimulus features coded by single neurons of a macaque body category selective patch. *Proceedings of the National Academy of Sciences of the United States of America*. 2016; 113:E2450–2459. <https://doi.org/10.1073/pnas.1520371113> PMID: 27071095
61. Premereur E, Taubert J, Janssen P, Vogels R, Vanduffel W. Effective connectivity reveals largely independent parallel networks of face and body patches. *Current Biology*. 2016; 26:3269–3279. <https://doi.org/10.1016/j.cub.2016.09.059> PMID: 27866893
62. Onken A, Liu JK, Karunasekara PPCR, Delis I, Gollisch T, Panzeri S. Using matrix and tensor factorizations for the single-trial analysis of population spike trains. *PLoS Comput Biol*. 2016; 12(11):1–46. <https://doi.org/10.1371/journal.pcbi.1005189> PMID: 27814363
63. Hyvärinen A, Gutmann M, Hoyer PO. Statistical model of natural stimuli predicts edge-like pooling of spatial frequency channels in V2. *BMC Neuroscience*. 2005; 6(12):1–12.
64. Hoyer PO, Hyvärinen A. A multi-layer sparse coding network learns contour coding from natural images. *Vision Research*. 2002; 42:1593–1605. PMID: 12074953
65. Hubel DH, Wiesel TN. Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*. 1959; 148:574–591. <https://doi.org/10.1113/jphysiol.1959.sp006308> PMID: 14403679
66. Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*. 1962; 160:106–154. <https://doi.org/10.1113/jphysiol.1962.sp006837> PMID: 14449617
67. Ringach DL. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology*. 2002; 88:455–463. <https://doi.org/10.1152/jn.2002.88.1.455> PMID: 12091567

68. Touryan J, Felsen G, Dan Y. Spatial structure of complex cell receptive fields measured with natural images. *Neuron*. 2005; 45:781–791. <https://doi.org/10.1016/j.neuron.2005.01.029> PMID: 15748852
69. Rehn M, Sommer FT. A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. *Journal of Computational Neuroscience*. 2007; 22:135–146. <https://doi.org/10.1007/s10827-006-0003-9> PMID: 17053994
70. Cowley BR, Smith MA, Kohn A, Yu BM. Stimulus-driven population activity patterns in macaque primary visual cortex. *PLoS Comput Biol*. 2017; 12(12):1–31.
71. Chichilnisky EJ. A simple white noise analysis of neuronal light responses. *Network*. 2001; 12:199–213. PMID: 11405422
72. Cowley BR, Smith MA, Kohn A, Yu BM. Stimulus-driven population activity patterns in macaque primary visual cortex. *PLoS Comput Biol*. 2016; 12(12): e1005185. <https://doi.org/10.1371/journal.pcbi.1005185> PMID: 27935935
73. Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, et al. Do we know what the early visual system does? *Journal of Neuroscience*. 2005; 25:10577–10597. <https://doi.org/10.1523/JNEUROSCI.3726-05.2005> PMID: 16291931
74. Turner MH, Giraldo LGS, Schwartz O, Rieke F. Stimulus- and goal-oriented frameworks for understanding natural vision. *Nature Neuroscience*. 2019; 22(1):15–24. <https://doi.org/10.1038/s41593-018-0284-0> PMID: 30531846
75. Orban GA. Higher order visual processing in macaque extrastriate cortex. *Physiological Review*. 2007; 88(1):59–89. <https://doi.org/10.1152/physrev.00008.2007> PMID: 18195083
76. Takahashi K, Gu Y, May PJ, Newlands SD, DeAngelis GC, Angelaki DE. Multimodal coding of three-dimensional rotation and translation in area MSTd: comparison of visual and vestibular selectivity. *Journal of Neuroscience*. 2007; 27:9742–9756. <https://doi.org/10.1523/JNEUROSCI.0817-07.2007> PMID: 17804635
77. Smith EC, Lewicki MS. Efficient auditory coding. *Nature*. 2006; 439(23):978–982.
78. Rokem A, Watzl S, Gollisch T, Stemmler M, Herz AV, Samengo I. Spike-timing precision underlies the coding efficiency of auditory receptor neurons. *Journal of Neurophysiology*. 2006; 95(4):2541–2552. <https://doi.org/10.1152/jn.00891.2005> PMID: 16354733
79. Hromádka T, DeWeese MR, Zador AM. Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol*. 2008; 6(1):e16. <https://doi.org/10.1371/journal.pbio.0060016> PMID: 18232737
80. Martinez CE, Goddard J, Persia LED, Milone DH, Rufiner HL. Denoising sound signals in a bioinspired non-negative spectro-temporal domain. *Digital Signal Processing*. 2015; 38:22–31.
81. David SV, Mesgarani N, Shamma SA. Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network: Computation in neural systems*. 2007; 18(3):191–212.
82. Leaver AM, Rauschecker JP. Cortical Representation of Natural Complex Sounds: Effects of Acoustic Features and Auditory Object Category. *Journal of Neuroscience*. 2010; 30(22):7604–7612. <https://doi.org/10.1523/JNEUROSCI.0296-10.2010> PMID: 20519535
83. Klein D, König P, Körding K. Sparse spectrotemporal coding of sounds. *EURASIP Journal on Applied Signal Processing*. 2003; 7:659–667.
84. Murthy VN. Olfactory maps in the brain. *Annual Review of Neuroscience*. 2011; 34:233–258. <https://doi.org/10.1146/annurev-neuro-061010-113738> PMID: 21692659
85. Broome BM, Jayaraman V, Laurent G. Encoding and decoding of overlapping odor sequences. *Neuron*. 2006; 51:467–482. <https://doi.org/10.1016/j.neuron.2006.07.018> PMID: 16908412
86. Koulakov AA, Rinberg D. Sparse Incomplete Representations: A Potential Role of Olfactory Granule Cells. *Neuron*. 2011; 72(1):124–136. <https://doi.org/10.1016/j.neuron.2011.07.031> PMID: 21982374
87. Gupta P, Albeanu DF, Bhalla US. Olfactory bulb coding of odors, mixtures and sniffs is a linear sum of odor time profiles. *Nature Neuroscience*. 2015; 18(2):272–281. <https://doi.org/10.1038/nn.3913> PMID: 25581362
88. Poo C, Isaacson JS. Odor representations in olfactory cortex: "sparse" coding, global inhibition, and oscillations. *Neuron*. 2009; 62(6):850–861. <https://doi.org/10.1016/j.neuron.2009.05.022> PMID: 19555653
89. Chen CFF, Zou DJ, Altomare CG, Xu L, Greer CA, Firestein SJ. Nonsensory target-dependent organization of piriform cortex. *Proceedings of the National Academy of Sciences*. 2014; 111(47):16931–16936.
90. Stettler DD, Axel R. Representations of odor in the piriform cortex. *Neuron*. 2009; 63(6):854–864. <https://doi.org/10.1016/j.neuron.2009.09.005> PMID: 19778513

91. Gottfried JA, Winston JS, Dolan RJ. Dissociable codes of odor quality and odorant structure in human piriform cortex. *Neuron*. 2006; 49(3):467–479. <https://doi.org/10.1016/j.neuron.2006.01.007> PMID: [16446149](https://pubmed.ncbi.nlm.nih.gov/16446149/)
92. Jadhav SP, Wolfe J, Feldman DE. Sparse temporal coding of elementary tactile features during active whisker sensation. *Nature Neuroscience*. 2009; 12:792–800. <https://doi.org/10.1038/nn.2328> PMID: [19430473](https://pubmed.ncbi.nlm.nih.gov/19430473/)
93. O'Connor DH, Peron SP, Huber D, Svoboda K. Neural activity in barrel cortex underlying vibrissa-based object localization in mice. *Neuron*. 2010; 67(6):1048–1061. <https://doi.org/10.1016/j.neuron.2010.08.026> PMID: [20869600](https://pubmed.ncbi.nlm.nih.gov/20869600/)
94. Crochet S, Poulet JFA, Kremer Y, Petersen CCH. Synaptic mechanisms underlying sparse coding of active touch. *Neuron*. 2011; 69(6):1160–1175. <https://doi.org/10.1016/j.neuron.2011.02.022> PMID: [21435560](https://pubmed.ncbi.nlm.nih.gov/21435560/)
95. Hafner W, Fend M, König P, Körding KP, Paul K. Predicting properties of the rat somatosensory system by sparse coding. *Neural information processing letters and reviews*. 2004; 4(1):11–18.
96. DiCarlo JJ, Johnson KO. Spatial and Temporal Structure of Receptive Fields in Primate Somatosensory Area 3b: Effects of Stimulus Scanning Direction and Orientation. *Journal of Neuroscience*. 2000; 20(1):495–510. PMID: [10627625](https://pubmed.ncbi.nlm.nih.gov/10627625/)
97. Bensmaia SJ, Denchev PV, Dammann JF, Craig JC, Hsiao SS. The Representation of Stimulus Orientation in the Early Stages of Somatosensory Processing. *Journal of Neuroscience*. 2008; 28(3):776–786. <https://doi.org/10.1523/JNEUROSCI.4162-07.2008> PMID: [18199777](https://pubmed.ncbi.nlm.nih.gov/18199777/)
98. Ramirez A, Pnevmatikakis EA, Merel J, Paninski L, Miller KD, Bruno RM. Spatiotemporal receptive fields of barrel cortex revealed by reverse correlation of synaptic input. *Nature Neuroscience*. 2014; 17(6):866–875. <https://doi.org/10.1038/nn.3720> PMID: [24836076](https://pubmed.ncbi.nlm.nih.gov/24836076/)
99. Pei YC, Hsiao SS, Craig JC, Bensmaia SJ. Neural mechanisms of tactile motion integration in somatosensory cortex. *Neuron*. 2011; 69:536–547. <https://doi.org/10.1016/j.neuron.2010.12.033> PMID: [21315263](https://pubmed.ncbi.nlm.nih.gov/21315263/)
100. Whiteway MR, Butts DA. Revealing unobserved factors underlying cortical activity with a rectified latent variable model applied to neural population recordings. *Journal of Neurophysiology*. 2017; 117:919–936. <https://doi.org/10.1152/jn.00698.2016> PMID: [27927786](https://pubmed.ncbi.nlm.nih.gov/27927786/)
101. Hyvärinen J, Poranen A. Receptive field integration and submodality convergence in the hand area of the post-central gyrus of the alert monkey. *Journal of Physiology*. 1978; 283:539–556. <https://doi.org/10.1113/jphysiol.1978.sp012518> PMID: [102768](https://pubmed.ncbi.nlm.nih.gov/102768/)
102. Pei YC, Denchev PV, Hsiao SS, Craig JC, Bensmaia SJ. Convergence of submodality-specific input onto neurons in primary somatosensory cortex. *Journal of Neurophysiology*. 2009; 102(3):1843–1853. <https://doi.org/10.1152/jn.00235.2009> PMID: [19535484](https://pubmed.ncbi.nlm.nih.gov/19535484/)
103. Harvey MA, Saal HP, III JFD, Bensmaia SJ. Multiplexing stimulus information through rate and temporal codes in primate somatosensory cortex. *PLoS Biol*. 2013; 11:1–11.
104. DiCarlo JJ, Johnson KO, Hsiao SS. Structure of receptive fields in area 3b of primary somatosensory cortex in the alert monkey. *Journal of Neuroscience*. 1998; 18(7):2626–2645. PMID: [9502821](https://pubmed.ncbi.nlm.nih.gov/9502821/)
105. London BM, Miller LE. Responses of somatosensory area 2 neurons to actively and passively generated limb movements. *Journal of Neurophysiology*. 2013; 109(6):1505–1513. <https://doi.org/10.1152/jn.00372.2012> PMID: [23274308](https://pubmed.ncbi.nlm.nih.gov/23274308/)
106. Miller AM, Vedder LC, Law LM, Smith DM. Cues, context, and long-term memory: the role of the retrosplenial cortex in spatial cognition. *Frontiers in human neuroscience*. 2014; 8:586. <https://doi.org/10.3389/fnhum.2014.00586> PMID: [25140141](https://pubmed.ncbi.nlm.nih.gov/25140141/)
107. Nelson AJD, Hindley E, Pearce JM, Vann SD, Aggleton JP. The effect of retrosplenial cortex lesions in rats on incidental and active spatial learning. *Frontiers in behavioral neuroscience*. 2015; 9:11. <https://doi.org/10.3389/fnbeh.2015.00011> PMID: [25705182](https://pubmed.ncbi.nlm.nih.gov/25705182/)
108. Vann SD, Aggleton JP, Maguire EA. What does the retrosplenial cortex do? *Nature Reviews Neuroscience*. 2009; 10(11):792–802. <https://doi.org/10.1038/nrn2733> PMID: [19812579](https://pubmed.ncbi.nlm.nih.gov/19812579/)
109. Alexander AS, Nitz DA. Retrosplenial cortex maps the conjunction of internal and external spaces. *Nature Neuroscience*. 2015; 18:1143–1151. <https://doi.org/10.1038/nn.4058> PMID: [26147532](https://pubmed.ncbi.nlm.nih.gov/26147532/)
110. Nelson AB, Kreitzer AC. Reassessing models of basal ganglia function and dysfunction. *Annual Review of Neuroscience*. 2014; 37:117–135. <https://doi.org/10.1146/annurev-neuro-071013-013916> PMID: [25032493](https://pubmed.ncbi.nlm.nih.gov/25032493/)
111. Lanciego JL, Luquin N, Obeso JA. Functional neuroanatomy of the basal ganglia. *Cold Spring Harbor Perspectives in Medicine*. 2012; 2:a009621. <https://doi.org/10.1101/cshperspect.a009621> PMID: [23071379](https://pubmed.ncbi.nlm.nih.gov/23071379/)

112. Schwab DJ, Houk JC. Presynaptic inhibition in the striatum of the basal ganglia improves pattern classification and thus promotes superior goal selection. *Frontiers in systems neuroscience*. 2015; 9:152. <https://doi.org/10.3389/fnsys.2015.00152> PMID: 26696840
113. Gillis N. The why and how of nonnegative matrix factorization. *arXiv [Preprint]*. 2014. 1401.5226.
114. Ding C, He X, Simon HD. On the equivalence of nonnegative matrix factorization and spectral clustering. In: *SIAM International Conference on Data Mining*; 2005. p. 606–610.
115. Ganguli S, Sompolinsky H. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annual Review of Neuroscience*. 2012; 35:485–508. <https://doi.org/10.1146/annurev-neuro-062111-150410> PMID: 22483042
116. Wang F, Li P. Compressed Nonnegative Sparse Coding. In: *2010 IEEE International Conference on Data Mining*; Piscataway, NJ: IEEE; 2010. p. 1103–1108.
117. Laughlin SB, Sejnowski TJ. Communication in neuronal networks. *Science*. 2003; 301(5641):1870–1874. <https://doi.org/10.1126/science.1089662> PMID: 14512617
118. Cherniak C. Component placement optimization in the brain. *Journal of Neuroscience*. 1994; 14(4):2418–2427. PMID: 8158278
119. Chklovskii DB, Schikorski T, Stevens CF. Wiring optimizations in cortical circuits. *Neuron*. 2002; 25(3):341–347.
120. Clopath C, Busing L, Vasilaki E, Gerstner W. Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nature Neuroscience*. 2010; 13:344–352. <https://doi.org/10.1038/nn.2479> PMID: 20098420
121. Rubinov M, Ypma RJF, Watson C, Bullmore ET. Wiring cost and topological participation of the mouse brain connectome. *Proceedings of the National Academy of Sciences of the United States of America*. 2015; 112(32):10032–10037. <https://doi.org/10.1073/pnas.1420315112> PMID: 26216962
122. Nessler B, Pfeiffer M, Maass W. STDP enables spiking neurons to detect hidden causes of their inputs. In: Bengio Y, Schuurmans D, Lafferty JD, Williams CKI, Culotta A, editors. *Advances in Neural Information Processing Systems 22*. Red Hook, NY: Curran Associates, Inc.; 2009. p. 1357–1365.
123. Carlson KD, Richert M, Dutt N, Krichmar JL. Biologically plausible models of homeostasis and STDP: Stability and learning in spiking neural networks. In: *The 2013 International Joint Conference on Neural Networks (IJCNN)*. Dallas, TX: IEEE; 2013. p. 1–8.
124. Oja E. Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*. 1982; 15(3):267–273. <https://doi.org/10.1007/BF00275687> PMID: 7153672
125. Froemke RC, Letzkus JJ, Kampa BM, Hang GB, Stuart GJ. Dendritic synapse location and neocortical spike-timing-dependent plasticity. *Frontiers in Synaptic Neuroscience*. 2010; 2:29. <https://doi.org/10.3389/fnsyn.2010.00029> PMID: 21423515
126. Bienenstock EL, Cooper LN, Munro PW. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*. 1982; 2(1):32–48. PMID: 7054394
127. Beyeler M, Carlson KD, Chou TS, Dutt ND, Krichmar JL. CARLsim 3: A user-friendly and highly optimized library for the creation of neurobiologically detailed spiking neural networks. In: *The 2015 International Joint Conference on Neural Networks (IJCNN)*. Killarney, Ireland: IEEE; 2015. p. 1–8.
128. Carlson KD, Nageswaran JM, Dutt N, Krichmar JL. An efficient automated parameter tuning framework for spiking neural networks. *Frontiers in Neuroscience*. 2014; 8:10. <https://doi.org/10.3389/fnins.2014.00010> PMID: 24550771
129. Crochet S, Petersen CCH. Correlating whisker behavior with membrane potential in barrel cortex of awake mice. *Nature Neuroscience*. 2006; 9:608–610. <https://doi.org/10.1038/nn1690> PMID: 16617340
130. Pachitariu M, Lyamzin DR, Sahani M, Lesica NA. State-dependent population coding in primary auditory cortex. *Journal of Neuroscience*. 2015; 35:2058–2073. <https://doi.org/10.1523/JNEUROSCI.3318-14.2015> PMID: 25653363
131. Musall S, Kaufman MT, Gluf S, Churchland AK. Movement-related activity dominates cortex during sensory-guided decision making. *bioRxiv [Preprint]*. 2018. <https://doi.org/10.1101/308288>
132. Mineault PJ, Tring E, Trachtenberg JT, Ringach DL. Enhanced spatial resolution during locomotion and heightened attention in mouse primary visual cortex. *Journal of Neuroscience*. 2016; 36:6382–6392. <https://doi.org/10.1523/JNEUROSCI.0430-16.2016> PMID: 27307228
133. Neill CM, Stryker MP. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*. 2010; 65:472–479. <https://doi.org/10.1016/j.neuron.2010.01.033> PMID: 20188652

134. Ayaz A, Saleem AB, Scholvinck ML, Carandini M. Layer-specific integration of locomotion and concurrent wall touching in mouse barrel cortex. *Current Biology*. 2013; 23:890–894. <https://doi.org/10.1016/j.cub.2013.04.012>
135. Saleem AB, Ayaz A, Jeffery KJ, Harris KD, Carandini M. Integration of visual motion and locomotion in mouse visual cortex. *Nature Neuroscience*. 2013; 16:1864–1869. <https://doi.org/10.1038/nn.3567> PMID: 24185423
136. Mazzucato L, Fontanini A, Camera GL. Stimuli reduce the dimensionality of cortical activity. *Frontiers in Systems Neuroscience*. 2016; 10:11. <https://doi.org/10.3389/fnsys.2016.00011> PMID: 26924968
137. Gallego JA, Perich MG, Miller LE, Solla SA. Neural manifolds for the control of movement. *Neuron*. 2017; 94:978–984. <https://doi.org/10.1016/j.neuron.2017.05.025> PMID: 28595054
138. Churchland MK, Shenoy KV. Temporal complexity and heterogeneity of single-neuron activity in pre-motor and motor cortex. *Journal of neurophysiology*. 2007; 97(6):4235–4257. <https://doi.org/10.1152/jn.00095.2007> PMID: 17376854
139. Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, et al. Demixed principal component analysis of neural population data. *Elife*. 2016; 5:e10989. <https://doi.org/10.7554/eLife.10989> PMID: 27067378
140. Mante V, Sussillo D, Shenoy KV, Newsome WT. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*. 2013; 503:78–84. <https://doi.org/10.1038/nature12742> PMID: 24201281
141. Tanaka H. Modeling the motor cortex: Optimality, recurrent neural networks, and spatial dynamics. *Neuroscience Research*. 2016; 104:64–71. <https://doi.org/10.1016/j.neures.2015.10.012> PMID: 26562334
142. Trautmann E, Stavisky S, Lahiri S, Ames K, Kaufman M, Ryu S, et al. Accurate estimation of neural population dynamics without spike sorting. *Neuron*. 2019. <https://doi.org/10.1016/j.neuron.2019.05.003> PMID: 31171448
143. Scott SH. Inconvenient truths about neural processing in primary motor cortex. *Journal of Physiology*. 2008; 586(5):1217–1224. <https://doi.org/10.1113/jphysiol.2007.146068> PMID: 18187462
144. Russo AA, Bittner SR, Perkins SM, Seely JS, London BM, Laura AH, et al. Motor cortex embeds muscle-like commands in an untangled population response. *Neuron*. 2018; 97(4):953–966. <https://doi.org/10.1016/j.neuron.2018.01.004> PMID: 29398358
145. Gallego JA, Perich MG, Naufel SN, Ethier C, Solla SA, Miller LE. Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nature Communications*. 2018; 9(1):4233. <https://doi.org/10.1038/s41467-018-06560-z> PMID: 30315158
146. Shenoy KV, Sahani M, Churchland MM. Cortical control of arm movements: a dynamical systems perspective. *Annual Review of Neuroscience*. 2013; 36:337–359. <https://doi.org/10.1146/annurev-neuro-062111-150509> PMID: 23725001
147. Fetz EE. Are movement parameters recognizably coded in the activity of single neurons? *Behavioral and Brain Sciences*. 1992; 15:679–690.
148. Scott SH. The role of primary motor cortex in goal-directed movements: insights from neurophysiological studies on non-human primates. *Current Opinion in Neurobiology*. 2003; 13:671–677. PMID: 14662367
149. Wu W, Hatsopoulos N. Evidence against a single coordinate system representation in the motor cortex. *Experimental Brain Research*. 2006; 175:197–210. <https://doi.org/10.1007/s00221-006-0556-x> PMID: 16775704
150. GoodSmith D, Chen X, Wang C, Kim SH, Song H, Burgalossi A, et al. Spatial representations of granule cells and mossy cells of the dentate gyrus. *Neuron*. 2017; 93:677–690. <https://doi.org/10.1016/j.neuron.2016.12.026> PMID: 28132828
151. Jung MW, McNaughton BL. Spatial selectivity of unit activity in the hippocampal granular layer. *Hippocampus*. 1993; 3:165–182.
152. Senzai Y, Buzsaki G. Physiological properties and behavioral correlates of hippocampal granule cells and mossy cells. *Neuron*. 2017; 93:691–704. <https://doi.org/10.1016/j.neuron.2016.12.011> PMID: 28132824
153. Marr D. Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society of London B, Biological Sciences*. 1971; 262(841):23–81. <https://doi.org/10.1098/rstb.1971.0078> PMID: 4399412
154. Treves A, Rolls ET. Computational analysis of the role of the hippocampus in memory. *Hippocampus*. 1994; 4:374–391. <https://doi.org/10.1002/hipo.450040319> PMID: 7842058
155. Yassa MA, Stark CEL. Pattern separation in the hippocampus. *Trends in Neurosciences*. 2011; 34:515–525. <https://doi.org/10.1016/j.tins.2011.06.006> PMID: 21788086

156. Knierim JJ, Neunuebel JP. Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics. *Neurobiology of Learning and Memory*. 2016; 129:38–49. <https://doi.org/10.1016/j.nlm.2015.10.008> PMID: 26514299
157. Skaggs WE, McNaughton BL, Wilson MA, Barnes CA. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*. 1996; 6:149–172. [https://doi.org/10.1002/\(SICI\)1098-1063\(1996\)6:2<149::AID-HIPO6>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1098-1063(1996)6:2<149::AID-HIPO6>3.0.CO;2-K) PMID: 8797016
158. Chen A, DeAngelis GC, Angelaki DE. Convergence of vestibular and visual self-motion signals in an area of the posterior sylvian fissure. *Journal of Neuroscience*. 2011; 31(32):11617–11627. <https://doi.org/10.1523/JNEUROSCI.1266-11.2011> PMID: 21832191
159. Poggio T. A theory of how the brain might work. *Cold Spring Harbor Symposium on Quantitative Biology*. 1990; 55:899–910.
160. Pouget A, Snyder LH. Computational approaches to sensorimotor transformations. *Nature Neuroscience*. 2000; 3(Suppl):1192–1198.
161. Nitz D. Parietal cortex, navigation, and the construction of arbitrary reference frames for spatial information. *Neurobiology of learning and memory*. 2009; 91(2):179–185. <https://doi.org/10.1016/j.nlm.2008.08.007> PMID: 18804545
162. Louie K, Glimcher PW, Webb R. Adaptive neural coding: from biological to behavioral decision-making. *Current opinion in behavioral sciences*. 2015; 5:91–99. <https://doi.org/10.1016/j.cobeha.2015.08.008> PMID: 26722666
163. Andersen RA, Snyder LH, Bradley DC, Xing J. Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience*. 1997; 20(1):303–330.
164. Rounds EL, Scott EO, Alexander AS, DeJong KA, Nitz DA, Krichmar JL. An Evolutionary Framework for Replicating Neurophysiological Data with Spiking Neural Networks. In: Handl J, Hart E, Lewis PR, López-Ibáñez M, Ochoa G, Paechter B, editors. *Parallel Problem Solving from Nature—PPSN XIV: 14th International Conference, Edinburgh, UK, September 17–21, 2016, Proceedings*. New York: Springer International Publishing; 2016. p. 537–547.
165. Fusi S, Miller EK, Rigotti M. Why neurons mix: high dimensionality for higher cognition. *Current opinion in neurobiology*. 2016; 37:66–74. <https://doi.org/10.1016/j.conb.2016.01.010> PMID: 26851755
166. Eichenbaum H. Barlow versus Hebb: When is it time to abandon the notion of feature detectors and adopt the cell assembly as the unit of cognition? *Neuroscience Letters*. 2017; 680:88–93. <https://doi.org/10.1016/j.neulet.2017.04.006> PMID: 28389238
167. Laughlin SB. Energy as a constraint on the coding and processing of sensory information. *Current Opinion in Neurobiology*. 2001; 11:475–480. PMID: 11502395