



Why neurons mix: high dimensionality for higher cognition

Stefano Fusi¹, Earl K Miller² and Mattia Rigotti³

Neurons often respond to diverse combinations of task-relevant variables. This form of mixed selectivity plays an important computational role which is related to the dimensionality of the neural representations: high-dimensional representations with mixed selectivity allow a simple linear readout to generate a huge number of different potential responses. In contrast, neural representations based on highly specialized neurons are low dimensional and they preclude a linear readout from generating several responses that depend on multiple task-relevant variables. Here we review the conceptual and theoretical framework that explains the importance of mixed selectivity and the experimental evidence that recorded neural representations are high-dimensional. We end by discussing the implications for the design of future experiments.

Addresses

¹ Center for Theoretical Neuroscience, Columbia University College of Physicians and Surgeons, USA

² The Picower Institute for Learning and Memory & Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA

³ IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA

Corresponding author: Fusi, Stefano (sf2237@columbia.edu)

Current Opinion in Neurobiology 2016, 37:66–74

This review comes from a themed issue on **Neurobiology of behavior**

Edited by **Alla Karpova** and **Roosbeh Kiani**

<http://dx.doi.org/10.1016/j.conb.2016.01.010>

0959-4388/Published by Elsevier Ltd.

Introduction

The traditional view of brain function is that individual neurons and even whole brain areas are akin to gears in a clock. Each is thought to be highly specialized for specific functions. This, however, does not fit with many observations, especially in higher-order cortex. For example, training monkeys on a cognitive-demanding task engages huge proportions of neurons in the prefrontal cortex (~40% of randomly sampled cells). This means that training either hijacks a huge slice of cortical tissue (and monkeys can only learn 2–3 tasks before their brains reach capacity). Or instead that neurons can do more than one thing. The latter does seem to be the case. Many neurons in the prefrontal and parietal cortices seem to be multitaskers.

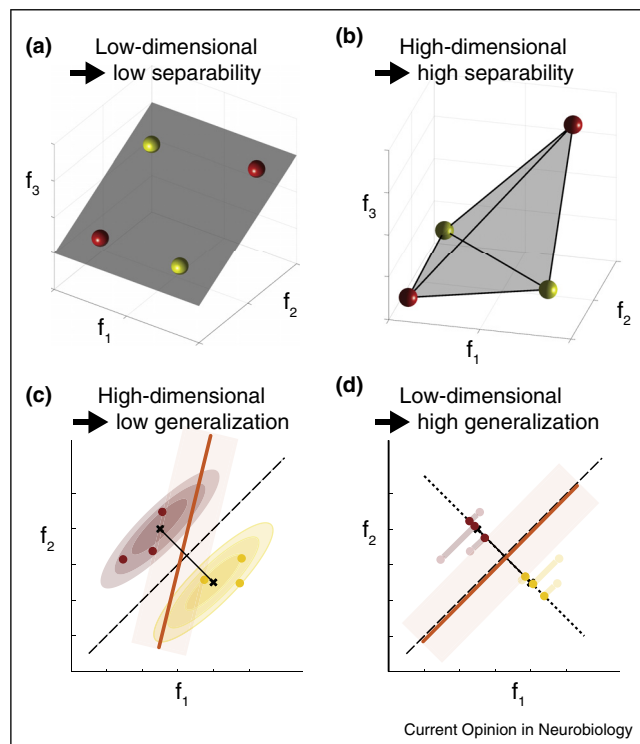
They behave differently in different contexts, as if they are members of different ensembles. This is a property we have termed ‘mixed selectivity’. Mixed selectivity neurons have been reported in a large body of experimental evidence, but only recent investigations have started to point out their possible importance for coding and the implementation of brain functions. Mixed selectivity can manifest itself as an ‘adaptive coding’ [1] of cortical cells whose responses are highly diverse and change over time. These responses encode multiple task-relevant variables that include rules, sensory stimuli identity or features, and motor responses or decisions [2–4,5^{••},6^{••},7–9,10[•],11[•],12[•]]. Mixed selectivity has also been reported in the hippocampus, where single units can respond to multiple contextual and episodic features [13–15,16^{••}], and in the amygdala, where neurons can be selective to specific combinations of visual stimuli, temporal context and predicted reinforcers during conditioning [4,17]). Why did the brain develop this unexpected property? Wouldn’t it be easier for each neuron or brain area to do one thing? It turns out, from a computational perspective, mixed selectivity may be central to complex behavior and cognition. A brain with neural representations based on highly specialized neurons would be hamstrung; only capable of learning a small number of simple tasks. Mixed selectivity endows the computational horsepower needed for complex thought and action. Here we summarize theoretical arguments developed in the computational neuroscience community that explain why. We then review the experimental evidence that supports the proposed interpretation of the computational role of mixed selectivity.

Understanding the computational role of mixed selectivity

Mixed selectivity neurons are selectively activated by combinations of different task variables that cannot be predicted by their responses to individual variables. These neuronal responses are actually repeatable: the neurons behave the same way in the same context, but their selectivity is highly context-dependent. As a consequence, the activity of any individual mixed selectivity neuron doesn’t mean anything by itself. Only in the context of other neurons it is possible to disambiguate the information encoded by mixed selectivity neurons. This fits with a recent update to the neuron doctrine notion, that ensembles, not individual neurons, are the functional unit of the nervous system [18].

However, encoding information is not enough. The information has to be explicit [19] so as to be accessible to downstream structures. Take, for example, the retina. All

Figure 1



(a), (b) Low and high-dimensional neural representations. The activity of a neuronal population of three neurons is represented as a point (visualized as a sphere) in the space of all possible patterns of activity. The three axes represent the firing rates f_k ($k = 1, 2, 3$) of the three neurons. The four spheres represent the population responses in four distinct experimental conditions (e.g. the responses to four sensory stimuli). The dimensionality of the neural representations is the minimal number of coordinate axes that are needed to specify the position of all points. (a) The points lie on a plane and hence they 'live in a low dimensional space' (2D). (b) A high-dimensional neural representation: same as in panel (a), but now the four points representing the sensory stimuli are no longer coplanar and they span three dimensions. This representation has the maximal dimensionality. In (a) a linear readout cannot be trained to separate the red from the yellow points as they all lie on a plane. This is because a linear readout can be trained only if there exists a plane (a hyperplane in higher dimensional spaces) that separates the red from the yellow points, which is clearly not the case here. This is a prototypical case of non-linear separability, and is equivalent to the well-known exclusive or (XOR) problem. It becomes possible to separate the yellow from the red points in (b), where the four points define a tetrahedron. As this geometrical arrangement gives the maximal dimensionality, all possible colorings of the three points are implementable by a linear readout. (c), (d) Dimensionality reduction can improve generalization. (c) Each shaded ellipse represents the distribution of response vectors in one of two specific conditions due to trial-to-trial noise. The centers of the clouds corresponding to the mean firing rates are on a line, but the points of the clouds are distributed across all two dimensions. In the example, the ellipses are elongated along the direction orthogonal to the black line that joins the centers of the clouds, indicating that noise is particularly high in that particular direction. Due to finite sampling, we might not be able to correctly estimate this noise structure, and this could result in a suboptimal readout. Say for instance that we were to train a linear readout only on the six points represented by the circles in the figure. In that case the resulting classifier (represented by the yellow separating line) would be clearly suboptimal with respect to the

the information needed for visual perception and recognition is there. However, the known circuits that are capable of reading it out in any useful way (such as the visual system) are overly complex. To determine if mixed selectivity representations are useful in terms of making information accessible to further processing stages, we need a yardstick to determine what sort of representations neural circuits can reasonably interpret. For this, we can turn to artificial neural networks. Simply put, if an artificial network based on simplifying biological principles can read out the relevant information, we assume the brain can too. A conservative measure would be a linear readout because it can be easily implemented as a weighted sum and threshold operation by individual units of an artificial network.

To understand the advantage of encoding information in a population of neurons with mixed selectivity to non-linear combinations of factors rather than a population of highly specialized neurons (what we'll call 'pure selectivity' neurons), consider Figure 1a. Each of the axes represents the firing rate of a different neuron, each one showing linear tuning to one factor or a linear combination of two factors. In other words, neurons without nonlinear mixed selectivity. Neuron 1 in the figure, whose firing rate is denoted by f_1 , is selective to sound in such a way that its activity increases linearly with sound intensity; neuron 2 is selective to visual inputs such that its activity increases linearly with visual contrast; the activity of neuron 3 is linearly related to either of those factors or a linear combination of the two factors (linear mixed selectivity). The four points represent the responses of the three neurons (response vectors) in four different conditions (meaning four different combinations of the factors). The task is to respond in one way to two of the combinations (shown in red) and in another way to the other two combinations (yellow). Because the neurons' firing has only a linear relationship to the two factors, these points are on a plane. A linear readout would need to find a plane that separates the task conditions of interest (red from yellow). But with linear neural tuning, one cannot find a linear readout that can separate yellow from red. One could find a readout that separates one larger factor from the rest, such as all conditions with loud sounds or low contrast, but it is not possible to separate different combinations of high and low signals.

overall distributions and would misclassify a considerable fraction of response vectors. A way to limit this finite sampling problem is to reduce dimensionality. In (d) the six points that were used to train the classifier are projected onto the dotted black line, the direction that discriminates between the two classes. Now, even with the limited sample of only six points it is possible to infer a separating hyperplane that would result in an optimal separation between the overall distributions.

? solved, see notes

Consider the case in which one of the neurons (Neuron 3) is a mixed selectivity neuron whose firing rate reflects a non-linear combination of the other two factors. The addition of just this one neuron makes the neural representations tridimensional as the four points are now the vertices of a tetrahedron. Now one can define a linear readout that can separate any arbitrary combination of the two factors. Using a different terminology taken from machine learning, the four points can be 'shattered' [20]. Or alternatively, all possible colorings (i.e. all ways of classifying the points as red and yellow) are implementable. More generally, the number of classifications that can be performed by a linear readout grows exponentially with the dimensionality of the neural representations of the inputs. As a consequence, a linear readout endowed with high-dimensional inputs can implement a much larger number of task-related responses. In general, high-dimensional representations require non-linear mixed selectivity of the kind illustrated in the previous example. Neural representations that are based on highly specialized neurons (pure selectivity neurons that are selective to an individual variable), or that respond to linear combinations of the task-relevant factors (linear mixed selectivity) are low dimensional as exemplified in Figure 1b. It is important to stress, that highly specialized neurons could be encoding all the relevant information about a task in segregated populations. However, these representations would be low dimensional, and therefore of limited use to a linear readout (see the Box for the upper bound on the number of dimensions).

Besides non-linear mixed selectivity, the second ingredient that is necessary for high dimensionality is diversity: different mixed selectivity neurons should exhibit different response properties. In other words, they should respond to different combinations of values of the task-relevant factors. The neural activation should be heterogeneous enough to guarantee a large coverage of the response space, generating what is technically known as a basis. A direct measure of the diversity of the neuronal responses has been proposed in [12^{*}], where the authors show quantitatively that neurons in the rat posterior parietal cortex (PPC) are highly heterogeneous when recorded while the animal performs a multi-sensory decision task.

Diversity and non-linear mixed selectivity are essential ingredients, but they might not be sufficient when noise is factored in. Noise can be defined as the component of the neuronal response that is not reproducible, as, for example, the fluctuations of the neural activity across different trials that correspond to repetitions of the same experimental condition. Because of noise, in any realistic situations the points in Figure 1a,b should actually be replaced by clouds of points, where each cloud represents the trial-to-trial distribution of the neural activity in a specific condition. Noise can introduce activity components that cause an increase in dimensionality. However,

Box How to determine the maximal dimensionality

If the number of neurons is larger than p , the number of points in firing rate space, then the maximal dimensionality attainable by the response vectors is equal to p . For example, if there are only two distinct points in firing rate space corresponding to two task conditions ($p = 2$), then there is always a line going through them, meaning that the dimensionality would be 1. To this we have to add a degree of freedom due to the offset from the origin. The total dimensionality is then 2, which is equal to the number of points. If the neurons are highly specialized and exhibit pure selectivity, then the dimensionality can be much smaller than p . To see this consider the case of two populations of pure selectivity neurons, one that encodes a task variable a and the other one that encodes a different variable b . If these variables have k values each, then the maximal dimensionality for each population is $k - 1$ (ignoring the offset). When these two populations are combined together all values of a and b can be combined in $p = k^2$ different ways but the maximal dimensionality would be $2k - 2$ plus the offset, which would give $2k - 1$. Clearly, for sufficiently large k , $2k - 1$ can be much smaller than k^2 . The gap becomes even larger when more than two variables are considered. For example, for n variables, the maximal dimensionality is k^n but the upper bound for neural representations with pure selectivity would be $nk - 1$. This indicates that the neural representations based on highly specialized pure selectivity neurons are significantly less efficient than those that incorporate non-linear mixed selectivity neurons when multiple variables have to be combined together. Also, it is important to note that the upper bound on the dimensionality depends on the complexity of the task, or more precisely on the number of distinct conditions. Simple tasks correspond to low dimensional neural representations, regardless of the number of recorded neurons (see also the discussion of [21]). There are situations in which it is not obvious to determine unambiguously the number of distinct conditions. For example, consider an experiment that involves a continuous task-relevant variable, like the direction of reaching in a motor task (see e.g. [22], one of the experiments analyzed in [21]). The animal is required to touch a target on a screen by executing reach movements in several possible directions. The target can appear at 27 different locations, corresponding to 27 discrete values of the reaching angle. One might then assume that the maximal dimensionality is 27. However, this postulates that the 27 conditions can be considered distinct task conditions. If this is not the case, because for instance directions corresponding to nearby angles are so similar to be barely distinguishable, then the number of reaching angles overestimates the maximal possible dimensionality on the response patterns. In [21] the authors formalized this problem and they essentially proposed that two conditions corresponding to nearby values of a task-relevant variable should be declared distinct only if the corresponding neuronal response vectors are distinguishable. That should then be the criterion defining in which sense conditions are to be considered distinct, which has to be used to bound the maximal dimensionality. This kind of analysis applied to the previously considered motor control experiment in the case where three consecutive reaching angles correspond to approximately the same neural response, would then reveal that the maximal dimensionality is approximately nine. The actual dimensionality could be even smaller if other correlations are present in the neural data. The formalism of [21] can be extended to continuous variables that are not discretized in the experiment, like time.

because of their non-repeatability, these components cannot be useful for coding and discriminate response vectors across conditions. For example, in Figure 1a the four points are on a plane and the task of separating the yellow from the red points cannot be performed by a

linear readout. When noise is introduced, the points will be displaced out of the plane, and for some noise realizations it might become possible to find a separating plane. However, this separating plane can only work for the specific noise realization to which it was (over)fit, and it will not solve the separability problem for other noise realizations. In general, noise components actually reduce the discriminability along their direction, even when the representations are high-dimensional as in the case of Figure 1b. Indeed, it is easy to imagine a situation in which the noise fluctuations are so large, that distinct conditions become difficult to discriminate (e.g. when the fluctuations are larger than the average distance between two clouds of points). In the section 'How to measure dimensionality' we discuss one possible way of determining dimensionality in the presence of noise.

High dimensional representations are not always desirable. Some situations actually require some form of dimensionality reduction. An example are classification tasks, which often benefit from representations that exclusively contain information that is relevant for discriminating among inputs in distinct classes, and hence are more robust to variations along unimportant input dimensions. We illustrate a simple example of dimensionality reduction in Figure 1c,d. where we consider two neurons, each responding to a different sound source (e.g. one to the sounds that arrive to the right ear and the other to the sounds that come to the left ear). Say that the task is to tell which sound is louder. The sounds are generated according to some distribution, which are reflected by the distributions of the firing rates of the two neurons represented by ovals in the figure. These distributions are chosen to make the task difficult, as the two sound intensities are often very hard to discriminate. If a linear readout is trained on a small number of sample sounds (those represented by spheres in the figure), it is possible that the resulting separating line is not the optimal one that would separate the two ovals and some of the future sounds will be misclassified. A better performance can be achieved by first projecting the data points on an appropriately chosen lower dimensional space, and then training the readout. In this case it would be desirable to project onto the line that unites the centers of the two distributions, reducing the dimensionality from 2D to 1D. This dimensionality reduction corresponds to the extraction of the relevant feature, which in this case is the difference in intensity between the two sounds. In short, the brain needs to reduce dimensionality to get rid of factors that are not relevant but at the same time recast the remaining factors into a high-dimensional space so that they can be processed to generate complex behavior.

How to measure dimensionality

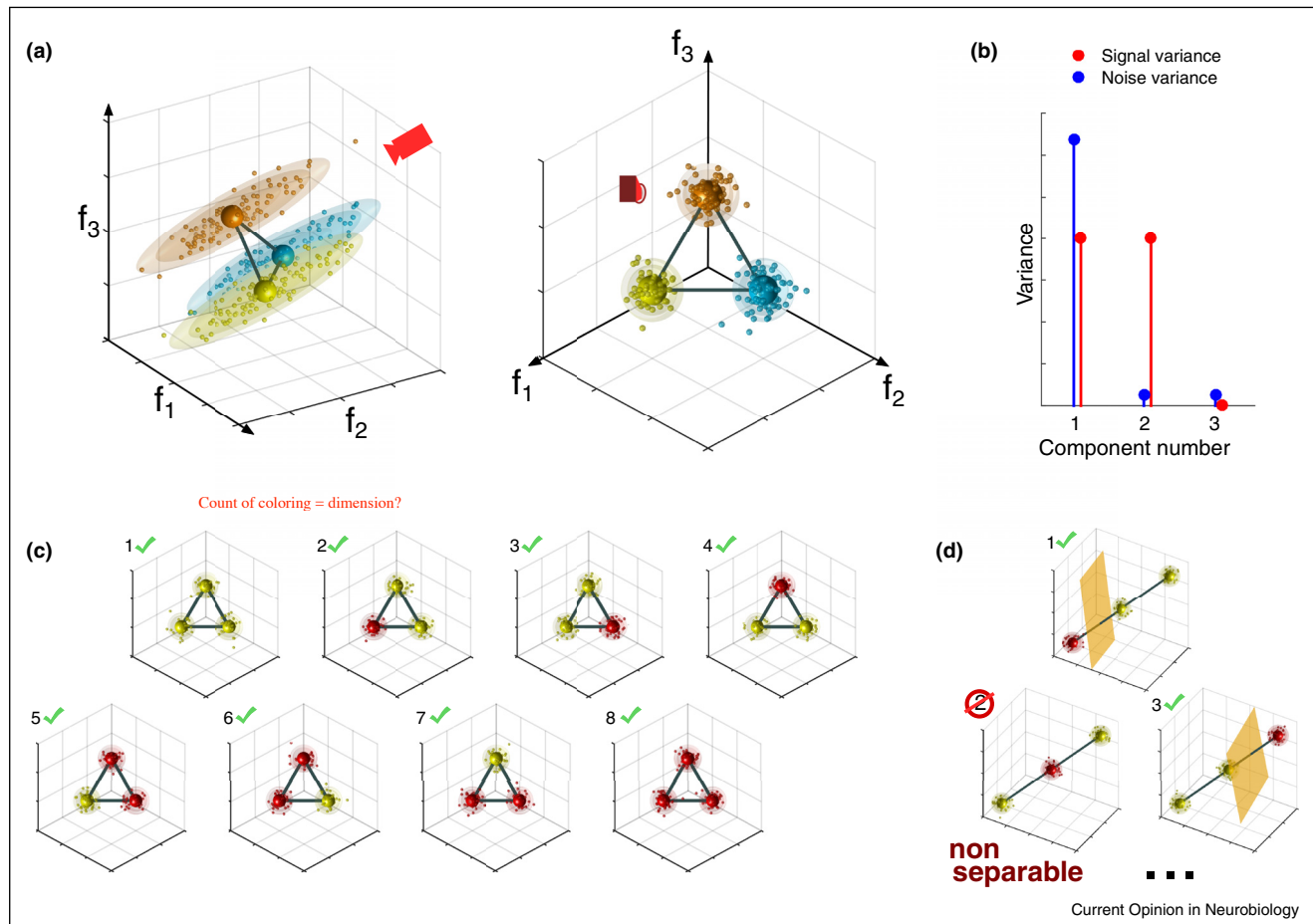
Dimensionality is the minimal number of coordinate axes needed to specify the positions of all points in firing rate space. In the case of Figure 1a, dimensionality is 2, as all

points are contained by a plane. In Figure 1b, dimensionality is 3, as the points are the vertices of a tetrahedron, a 3D object. More formally, dimensionality can be defined as the rank of the matrix F whose columns represent the firing rate vectors. The number of rows would be equal to the number of neurons, and the number of columns are equal to the number of conditions. The practical problem with this definition is finite sampling noise, since any imprecision in estimating the actual firing rates by averaging out trial-to-trial fluctuations in the activity artificially inflates dimensionality estimates. For example, in the Figure 1a, the noise would pull the points out from the plane and the rank of F would be maximal (as large as the number of conditions, assuming that the number of recorded neurons is larger), making us unable to discriminate between the geometry of Figure 1a (low dimensional) and the one of Figure 1b (high-dimensional).

A useful method of estimating dimensionality that takes noise into account has been recently developed utilizing techniques related to Principal Component Analysis (PCA) [23]. PCA identifies the directions along which the points in the firing rate space have the broadest distributions (variance). These directions are called principal components and are characterized by the amount of variance in the data that can be captured or 'explained' by them. For example, in Figure 1a, two components will be the axes that define the plane, and the third one would be orthogonal to it. The variance will be large for the first two components, and zero for the third one. So the dimensionality, which is 2, is simply the number of principal components with a non-zero variance. If a small amount of noise is added to the firing rates, then the third component will be only slightly different from zero, and one could still estimate the right dimensionality for moderate and known noise magnitudes by cutting off the smallest components. More formally, the eigenvectors of the covariance matrix FF^T are the principal components, and the corresponding eigenvalues determine the variance that is explained by each component. In the noiseless case, the number of non-zero eigenvalues is equal to the rank of the matrix F . In the noisy case, the dimensionality can be defined as the number of eigenvalues that are above some threshold. The noise threshold can be estimated from the trial-to-trial fluctuations of the neuronal responses [23].

Notice that this method allows us to count all dimensions along which there is some variation, whether they are useful or not for the task. For example, in Figure 2a,b, the strongest principal component is actually orthogonal to the line along which the three clouds of points can be separated. Moreover, this method can be inaccurate as it usually provides a lower bound on the dimensionality, given that the threshold is estimated on the basis of the largest noise component.

Figure 2



Measuring the dimensionality of neural representations: the effects of noise and correlation on different estimation methods. **(a)** The plot depicts the responses of three neurons (f_1 , f_2 , f_3) to three task conditions (red, green, blue). The 3D coordinates of each dot correspond to the joint response of the three neurons in an individual trial. Color corresponds to the condition eliciting the response. The spheres indicate the mean trial-averaged activity in the three conditions. Panel one: example neurons display high positive pairwise noise correlation (represented by the transparent ellipsoids) between neurons that results in a large trial-to-trial variability along the direction pointing to the positive orthant for all three conditions. Panel two: the same data of panel one from a different angle in the 3D activity space. This shows that, along appropriately chosen directions, the response vectors in the three conditions can be separated from each other with a very high signal-to-noise ratio. The correlated noise along the (1,1,1)-direction does not limit the discrimination between population response patterns, since it is orthogonal to the directions separating the trial-averaged activity patterns. (See e.g. [58–60] for more general consideration regarding when and how noise correlations can help or harm signal decoding.) **(b)** Looking at the absolute variance of the signal and noise separately in the example in (a) might give the misleading impression that the correlated variability ablates all coding information, since for instance a PCA analysis (see e.g. [23]) shows that the noise component is much higher than the signal in the two coding directions. In fact, since the main noise component lies along a direction that is orthogonal to the coding directions, it does not have any influence on decoding the condition from the population response patterns. **(c)** Use of population decoding to characterize response patterns dimensionality. Despite the high noise (panel (b)) an appropriate decision boundary can discriminate any arbitrary binary partition of the task conditions (in the case of three conditions there exist $2^3 = 8$ such partitions). This means that a linear decoder can be trained to implement any possible binary function over task conditions using the corresponding response patterns as inputs. In addition, this is related to the dimensionality of the vector space effectively spanned by the trial-averaged response patterns: the three average within-condition response patterns lie on a 2D space. Because this dimensionality is exponentially related to the number of binary functions that can be implemented by a linear decoder, it can serve as a measure of the number of responses that can be reliably implemented by downstream circuits that are at least as complex as a linear readout. **(d)** Example where the same number of average response patterns effectively lie on a lower dimensional space (in this case a line). In such a situation, not all binary partitions of inputs can be implemented by a linear classifier, as indicated by the non-separable partitioning of patterns in the second plot.

An alternative method that is more directly related to the computational advantages of dimensionality has been employed in [6**]. The fact that apparent increases in dimensionality due to noise components allow for

separations that do not generalize to different noise realizations suggests that dimensionality can be estimated by counting the number of linearly separable colorings of the firing rate vectors through cross-validation. The idea

is that separations that exploit non-repeatable noise components will be revealed as artifactual if one **tests them on different realizations of noise** (corresponding to different trials). This means that we can estimate dimensionality by counting the number of **linearly separable colorings of the firing rate vectors that maintain a high cross-validation performance across noise realizations** (see Figure 2). An advantage of this method is that it actually takes into account the shape and orientation of the noise distributions in relation to the coding directions that separate the points. The technique in [23] assumes a worst-case scenario where the largest noise component is aligned to the coding dimensions that are at most as large as the noise (e.g. in Figure 2c, **dimensionality would be 1, and not 2 as for the PCA analysis**).

Dimensionality transformations

Quantifying the dimensionality of the neural representations offers an informative glimpse on **how sensory and cognitive signals are being processed in the recorded brain areas**. This observation derives from recent computational studies aimed at understanding neural coding in sensory and higher-order cortical areas through the lens of modern machine learning methods such as deep artificial neural networks [24]. One of the fundamental insights gained from such an interdisciplinary effort is that the solution of tasks that require processing of natural sensory stimuli (natural images, speech, language) **benefit from a decomposition in multiple hierarchical elaboration stages of increasing level of abstraction**, a notion that has clear roots in the pioneering work of Hubel and Wiesel [25]. The ultimate goal of such stagewise processing is to **encode the relation between high-dimensional sensory inputs and low-dimensional decisions variables**.

A central strategy employed by successful deep neural networks is to strike a balance between dimensionality expansion and dimensionality reduction as processing progresses throughout the hierarchy.

Dimensionality reduction in sensory processing

Dimensionality reduction (obtained for instance by **pooling among competing input units guarantees better generalization by enforcing invariance to input changes that should not affect the final output** [26–30]). This is important in object recognition, where it is desirable that **different views of the same object (say, due to slight changes in pose, position, and so on) all result in the same classification response**. Geometrically this can be conceptualized by a transformation that collapses a large set of high-dimensional retinal inputs all corresponding to different variations of an object to the same high-level representation (see Figure 1c,d). **The resulting representation clearly allows for a potentially large reduction in sample complexity**, since the label provided by a training sample can be generalized over all the inputs corresponding to the same object [26,28].

Invariance computation has been proposed as an algorithmic principle for object recognition and a hallmark of the primate visual system [19,27], and is consistent with the selectivity to highly specific visual percept of neurons famously reported in inferotemporal (IT) cortex [31]. What's more, deep neural networks that are optimized to perform object recognition have been recently shown to spontaneously generate in their higher stages of processing neural activations that are highly predictive of V4 and IT cortex responses [32**].

Dimensionality expansion for input separability and output flexibility

As we pointed out in our first example, the opposite operation to dimensionality reduction, dimensionality expansion, is **important for cognition if one is to perform tasks that involve decisions over multiple variables** (see Figure 1a,b). Dimensionality expansion is one of the fundamental operations performed by modern machine learning algorithms (Support Vector Machines [33]) and by deep artificial neural networks [24]. It has the function of guaranteeing **high-margin separation and discrimination between inputs that have to generate distinct responses** (see e.g. [6**,34]). Moreover, high-dimensional neural representations are also a crucial component of dynamical models of **recurrent neural networks like echo state machines or liquid state machines** [35–40]. These models exhibit a very rich behavior similar to the one observed in cortical recordings and can solve complex tasks that require some form of short term or working memory. They do that by **non-linearly mixing the current state of a recurrent network with the external input**, generating **high-dimensional neural representations that are continuously updated with information about recent sensory stimuli**. Dimensionality is often expanded as a result of an iterative procedure that trains intermediate layers of hidden neurons (e.g. with back-propagation [41]). It can also be achieved by engineering neurons that display mixed selectivity simply by choosing the proper synaptic weights (if the tuning curves in the inputs are known). For example, mixed selectivity to the retinal location of a visual stimulus and the position of the eyes can be generated by gain fields that represent the stimulus position in order to determine reach movements in joint coordinates [42–44]. This kind of mixed selectivity to stimulus identity and to a context signal has been used to model visuomotor remapping [45–48].

An alternative and surprisingly general approach for generating high-dimensional neural representations is to **introduce layers of non-linear neurons that have random synaptic weights**. Random projections are extremely efficient at generating mixed selectivity and expand dimensionality (e.g. in [35–39,49]), without compromising the ability to generalize [50,51]. The analysis of the patterns of connectivity between glomeruli and Kenyon cells revealed that random projections could be the

strategy to expand dimensionality in the olfactory system of the drosophila [52[•]]. Interestingly dimensionality in sensory areas is a quantity that recent imaging studies suggest being preserved across individuals. Indeed, it is known from fMRI and electrophysiological studies that the dissimilarity matrix, which essentially contains the distances between the points in the firing rate space and hence determines the dimensionality is often invariant across subjects (hyperalignment) [53], and across species [54].

Measures of dimensionality and their correlation with behavior

Recent experimental recordings in primates during the execution of a cognitive working memory task showed that the dimensionality of the neural representations in pre-frontal cortex reaches the maximal value accommodated by the stimulus-response space [6^{••}]. Crucially, both in rodents as well as in primates it was also observed that dimensionality in frontal areas collapses on error trials, suggesting that the high dimensionality of the recorded representations is important for correct execution of the task [6^{••},55]. Interestingly, pharmacological intervention has also been recently shown to affect dimensionality. Specifically, moderate dosage of amphetamine enhanced the separation of high-dimensional components of PFC activity. This effect was however reversed at higher concentrations [56].

The collapse of dimensionality preceding errors could have at least two possible explanations: **the animals make a mistake because of a failure of perception** or memory (e.g. when one of the sensory cues is not seen, or one is forgotten) or the error is due to **a more subtle problem forming the mixed combinations of information**. It was found in [6^{••}] that the second explanation seems to be correct, as the identity of the two visual cues could still be decoded with high accuracy on error trials. This suggests that mixed selectivity neurons play an important role in representing the information about the visual cues in a specific high dimensional format that can be utilized by downstream circuits to generate behavior. This hypothesis then predicts that **a collapse in dimensionality impairs the ability of downstream readout neurons to produce a correct response**, explaining the correlation between a drop in dimensionality and behavioral errors. Finally, it was shown that the collapse in dimensionality has to be ascribed to a change in the statistics of the non-linear mixed selectivity component of the neuronal responses [6^{••}].

Conclusions

The cortex is not just a patchwork of highly specialized cells, each neuron is pretty much unique in its response properties and there is a good computational reason for this diversity: neural representations based on pure selectivity greatly limit the type of responses that can be

implemented by downstream readouts. However, it is worth stressing that there are situations in which highly specialized cells can be beneficial. For example when the task depends only on a single variable, then the best representations are those that encode just that variable. Given the diversity of the neurons, it is then important to consider and analyze their activity at the population level. In particular, in the search for the brain areas that are involved in a particular task, one should consider not only whether the task-relevant information is represented, but also in what way it is represented. In tasks in which the decisions of the subject depend on multiple variables, we predict that the brain areas that are causally linked to behavior are those that exhibit highest dimensionality. In experiments aimed at testing this hypothesis it would be important to make the task sufficiently complex so that dimensionality varies in a wide enough range to easily detect changes. The range is essentially determined by the maximal dimensionality, which depends on both the complexity of the task (essentially the number of independent conditions of the experiment, see the Box for more details) and on the correlations between neural representations. It would be advisable to make the task as complex as possible, which means having a large number of different conditions, but still compatible with the ability to repeat each condition a sufficient number of times to collect enough statistics about the recorded activity.

Dimensionality is a global property of the geometry of the neural representations, and it can be studied also using fMRI multi voxel pattern analysis (see e.g. [53]). Studying brain areas with high-dimensional representations and no anatomical organization, which would be rich of mixed selectivity neurons, might require techniques like repetition suppression to determine accurately the dimensionality of the neural representations [57].

Conflict of interest statement

Nothing declared.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Duncan J: **The structure of cognition: attentional episodes in mind and brain.** *Neuron* 2013, **80**:35-50 <http://dx.doi.org/10.1016/j.neuron.2013.09.015>.
2. Mansouri FA, Matsumoto K, Tanaka K: **Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin card sorting test analog.** *J Neurosci* 2006, **26**:2745-2756.
3. Churchland MM, Shenoy KV: **Temporal complexity and heterogeneity of single-neuron activity in premotor and motor cortex.** *J Neurophysiol* 2007, **97**:4235-4257.

4. Rigotti M, Rubin DBD, Morrison SE, Salzman CD, Fusi S: **Attractor concretion as a mechanism for the formation of context representations.** *Neuroimage* 2010, **52**:833-847.
5. Mante V, Sussillo D, Shenoy KV, Newsome WT: **Context-dependent computation by recurrent dynamics in prefrontal cortex.** *Nature* 2013, **503**:78-84 <http://dx.doi.org/10.1038/nature12742>.
- Monkeys are trained to integrate different features of noisy sensory stimuli. The relevant feature depends on the context, which varies across trials. The authors show that neurons exhibit very diverse responses and are selective to both the relevant and the irrelevant feature. The observed complexity of the individual neuronal responses is explained in the framework of a high-dimensional dynamical process at the level of the population
6. Rigotti M, Barak O, Warden MR, Wang X-J, Daw ND, Miller EK, Fusi S: **The importance of mixed selectivity in complex cognitive tasks.** *Nature* 2013, **497**:585-590 <http://dx.doi.org/10.1038/nature12160>.
- Mixed selectivity in single-cell recordings in the prefrontal cortex of primates during a memory task is shown to be a signature of a high-dimensional population encoding of task-relevant information. Changes in the dimensionality of the neural population activity patterns were moreover predictive of animal behavior
7. Jun JK, Miller P, Hernández A, Zainos A, Lemus L, Brody CD, Romo R: **Heterogeneous population coding of a short-term memory and decision task.** *J Neurosci* 2010, **30**:916-929 <http://dx.doi.org/10.1523/JNEUROSCI.2062-09.2010>.
8. Barak O, Tsodyks M, Romo R: **Neuronal population coding of parametric working memory.** *J Neurosci* 2010, **30**:9424-9430 <http://dx.doi.org/10.1523/JNEUROSCI.1875-10.2010>.
9. Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J: **Dynamic coding for cognitive control in prefrontal cortex.** *Neuron* 2013, **78**:364-375 <http://dx.doi.org/10.1016/j.neuron.2013.01.039>.
10. Pagan M, Urban LS, Wohl MP, Rust NC: **Signals in inferotemporal and perirhinal cortex suggest an untangling of visual target information.** *Nat Neurosci* 2013, **16**:1132-1139.
- The authors recorded the neural activity in inferotemporal cortex (IT) and perirhinal cortex (PRH) of a monkey performing a task that required to find targets in sequences of distractors. They found that the neural representations in the two areas contained similar amounts of task-specific information. However, the information was more accessible using a linear read-out or, equivalently, was higher dimensional in PRH.
11. Meister ML, Hennig JA, Huk AC: **Signal multiplexing and single-neuron computations in lateral intraparietal area during decision-making.** *J Neurosci* 2013, **33**:2254-2267. Neurons in primate lateral intraparietal cortex were recorded during a visual direction-discrimination task, while manipulating the presence of a saccade target in their receptive fields during decision-making. Cells were shown to multiplex decision signals with decision-irrelevant visual signals, and also display heterogeneous dynamics that is much more complex than the conventionally reported ramping activity..
12. Raposo D, Kaufman MT, Churchland AK: **A category-free neural population supports evolving demands during decision-making.** *Nat Neurosci* 2014.
- Novel experimental protocols and data analysis methods are developed for the study of multisensory decisions in rats. Neural activations in posterior parietal cortex showed a random distribution of selectivities for task-related parameters, and a reshaping of neural covariance structure in different task epochs, suggesting that population responses dynamically support evolving behavioral demands.
13. McNaughton BL, Barnes CA, O'Keefe J: **The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats.** *Exp Brain Res* 1983, **52**:41-49.
14. Markus EJ, Qin YL, Leonard B, Skaggs WE, McNaughton BL, Barnes CA: **Interactions between location and task affect the spatial and directional firing of hippocampal neurons.** *J Neurosci* 1995, **15**:7079-7094 <http://markus.lab.uconn.edu/wp-content/uploads/sites/1005/2015/01/Markus1995.pdf>.
15. Wood ER, Dudchenko PA, Robitsek RJ, Eichenbaum H: **Hippocampal neurons encode information about different types of memory episodes occurring in the same location.** *Neuron* 2000, **27**:623-633 <http://hsinnamon.web.wesleyan.edu/wescourses/NSB-Psyc255/Readings/17>.
16. McKenzie S, Frank AJ, Kinsky NR, Porter B, Rivière PD, Eichenbaum H: **Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas.** *Neuron* 2014, **83**:202-215 <http://dx.doi.org/10.1016/j.neuron.2014.05.019>.
- The responses of hippocampal neurons are very diverse and encode all task relevant variables (context, position, reward, object identity). By analyzing populations responses, the authors show that the neurons do not respond to a random combination of these variables, but the hippocampal representations exhibit a hierarchical organization.
17. Saez A, Rigotti M, Ostojic S, Fusi S, Salzman CD: **Abstract context representations in primate amygdala and prefrontal cortex.** *Neuron* 2015, **87**:869-881 <http://dx.doi.org/10.1016/j.neuron.2015.07.024>.
18. Yuste R: **From the neuron doctrine to neural networks.** *Nat Rev Neurosci* 2015, **16**:487-497.
19. DiCarlo JJ, Zoccolan D, Rust NC: **How does the brain solve visual object recognition?** *Neuron* 2012, **73**:415-434.
20. Vapnik VN: *The nature of statistical learning theory. Statistics for engineering and information science.* New York: Springer-Verlag; 2000.
21. Gao P, Ganguli S: **On simplicity and complexity in the brave new world of large-scale neuroscience.** *Curr Opin Neurobiol* 2015, **32**:148-155.
22. Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, Shenoy KV: **Neural population dynamics during reaching.** *Nature* 2012, **487**:51-56.
23. Machens CK, Romo R, Brody CD: **Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex.** *J Neurosci* 2010, **30**:350-360 <http://dx.doi.org/10.1523/JNEUROSCI.3276-09.2010>.
24. LeCun Y, Bengio Y, Hinton G: **Deep learning.** *Nature* 2015, **521**:436-444 <http://dx.doi.org/10.1038/nature14539>.
25. Hubel DH, Wiesel TN: **Receptive fields of single neurones in the cat's striate cortex.** *J Physiol* 1959, **148**:574-591.
26. Riesenhuber M, Poggio T: **Hierarchical models of object recognition in cortex.** *Nat Neurosci* 1999, **2**:1019-1025.
27. DiCarlo JJ, Cox DD: **Untangling invariant object recognition.** *Trends Cogn Sci* 2007, **11**:333-341.
28. Anselmi F, Rosasco L, Poggio T: *On invariance and selectivity in representation learning.* 2015arXiv:1503.05938.
29. Wang W, Carreira-Perpinán MA: **The role of dimensionality reduction in classification.** *Proc. of the 28th National Conference on Artificial Intelligence (AAAI 2014); Quebec City, Canada: 2014.*
30. Cadieu CF, Hong H, Yamins D, Pinto N, Majaj NJ, DiCarlo JJ: *The neural representation benchmark and its evaluation on brain and machine.* 2013arXiv:1301.3530.
31. Gross CG, Rocha-Miranda C, Bender D: **Visual properties of neurons in inferotemporal cortex of the macaque.** *J Neurophysiol* 1972, **35**:96-111.
32. Yamins DL, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ: **Performance-optimized hierarchical models predict neural responses in higher visual cortex.** *Proc Natl Acad Sci U S A* 2014, **111**:8619-8624.
- Artificial neurons of deep artificial neural network models optimized to perform object classification are compared to the responses to natural images of neurons in monkeys. The neural activations in the higher layers of the artificial network became highly predictive of the responses of neurons in the primate V4 and interior temporal cortex.
33. Cortes C, Vapnik V: **Support-vector networks.** *Mach Learn* 1995, **20**:273-297 <http://dx.doi.org/10.1007/BF00994018>.
34. Barak O, Rigotti M: **A simple derivation of a bound on the perceptron margin using singular value decomposition.** *Neural Comput* 2011, **23**:1935-1943 http://dx.doi.org/10.1162/NECO_a_00152.

35. Jaeger H, Haas H: **Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication.** *Science* 2004, **304**:78-80 <http://dx.doi.org/10.1126/science.1091277>.
36. Maass W, Natschläger T, Markram H: **Real-time computing without stable states: a new framework for neural computation based on perturbations.** *Neural Comput* 2002, **14**:2531-2560 <http://dx.doi.org/10.1162/089976602760407955>.
37. Buonomano DV, Maass W: **State-dependent computations: spatiotemporal processing in cortical networks.** *Nat Rev Neurosci* 2009, **10**:113-125 <http://dx.doi.org/10.1038/nrn2558>.
38. Sussillo D, Abbott LF: **Generating coherent patterns of activity from chaotic neural networks.** *Neuron* 2009, **63**:544 <http://dx.doi.org/10.1016/j.neuron.2009.07.018>.
39. Rigotti M, Ben Dayan Rubin D, Wang X-J, Fusi S: **Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses.** *Front Comput Neurosci* 2010, **4**:29 <http://dx.doi.org/10.3389/fncom.2010.00024>.
40. Laje R, Buonomano DV: **Robust timing and motor patterns by taming chaos in recurrent neural networks.** *Nat Neurosci* 2013, **16**:925-933.
41. Zipser D, Andersen RA: **A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons.** *Nature* 1988, **331**:679-684.
42. Andersen RA, Zipser D: **The role of the posterior parietal cortex in coordinate transformations for visual-motor integration.** *Can J Physiol Pharmacol* 1988, **66**:488-501.
43. Pouget A, Sejnowski TJ: **Spatial transformations in the parietal cortex using basis functions.** *J Cogn Neurosci* 1997, **9**:222-237.
44. Salinas E, Abbott L: **Coordinate transformations in the visual system: how to generate gain fields and what to compute with them.** *Prog Brain Res* 2001, **130**:175-190.
45. Salinas E: **Fast remapping of sensory stimuli onto motor actions on the basis of contextual modulation.** *J Neurosci* 2004, **24**:1113-1118.
46. Bourjaily MA, Miller P: **Synaptic plasticity and connectivity requirements to produce stimulus-pair specific responses in recurrent networks of spiking neurons.** *PLoS Comput Biol* 2011, **7**:e1001091 <http://dx.doi.org/10.1371/journal.pcbi.1001091>.
47. Bourjaily MA, Miller P: **Dynamic afferent synapses to decision-making networks improve performance in tasks requiring stimulus associations and discriminations.** *J Neurophysiol* 2012, **108**:513-527.
48. Rutishauser U, Douglas RJ: **State-dependent computation using coupled recurrent networks.** *Neural Comput* 2009, **21**:478-509.
49. Huang G-B, Zhu Q-Y, Siew C-K: **Extreme learning machine: theory and applications.** *Neurocomputing* 2006, **70**:489-501.
50. Barak O, Rigotti M, Fusi S: **The sparseness of mixed selectivity neurons controls the generalization-discrimination trade off.** *J Neurosci* 2013, **33**:3844-3856 <http://dx.doi.org/10.1523/JNEUROSCI.2753-12.2013>.
51. Babadi B, Sompolinsky H: **Sparseness and expansion in sensory representations.** *Neuron* 2014, **83**:1213-1226.
52. Caron SJ, Ruta V, Abbott L, Axel R: **Random convergence of olfactory inputs in the drosophila mushroom body.** *Nature* 2013, **497**:113-117 Using anatomical tracing techniques, the authors reconstructed the connectivity between glomeruli and Kenyon cells in the Drosophila. They show that each Kenyon cell integrates input from an apparently random combination of glomeruli. These random connections could be used to expand the dimensionality of the glomerular neural representations in order to facilitate learning of olfactory associations and behaviors..
53. Haxby JV, Guntupalli JS, Connolly AC, Halchenko YO, Conroy BR, Gobbini MI, Hanke M, Ramadge PJ: **A common, high-dimensional model of the representational space in human ventral temporal cortex.** *Neuron* 2011, **72**:404-416.
54. Lehky SR, Kiani R, Esteky H, Tanaka K: **Dimensionality of object representations in monkey inferotemporal cortex.** *Neural Comput* 2014.
55. Balaguer-Ballester E, Lapish CC, Seamans JK, Durstewitz D: **Attracting dynamics of frontal cortex ensembles during memory-guided decision-making.** *PLoS Comput Biol* 2011, **7**:e1002057 <http://dx.doi.org/10.1371/journal.pcbi.1002057>.
56. Lapish CC, Balaguer-Ballester E, Seamans JK, Phillips AG, Durstewitz D: **Amphetamine exerts dose-dependent changes in prefrontal cortex attractor dynamics during working memory.** *J Neurosci* 2015, **35**:10172-10187 <http://dx.doi.org/10.1523/JNEUROSCI.2421-14.2015>.
57. Rigotti M, Fusi S: **Estimating the dimensionality of neural responses with fMRI repetition suppression.** *MLINI workshop at NIPS*. 2015.
58. Abbott LF, Dayan P: **The effect of correlated variability on the accuracy of a population code.** *Neural Comput* 1999, **11**:91-101.
59. Averbeck BB, Latham PE, Pouget A: **Neural correlations, population coding and computation.** *Nat Rev Neurosci* 2006, **7**:358-366 <http://dx.doi.org/10.1038/nrn1888>.
60. Moreno-Bote R, Beck J, Kanitscheider I, Pitkow X, Latham P, Pouget A: **Information-limiting correlations.** *Nat Neurosci* 2014, **17**:1410-1417 <http://dx.doi.org/10.1038/nn.3807>.