Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Neurobiology

Why neurons mix: high dimensionality for higher cognition

Stefano Fusi[1], Earl K Miller[2] and Mattia Rigotti[3]

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 cognitive behavior**

Edited by **Alla Karpova** and **Roozbeh Kiani**

For a complete overview see the Issue and the Editorial

Available online 4th February 2016

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

0959-4388/Published by Elsevier Ltd.

Introduction

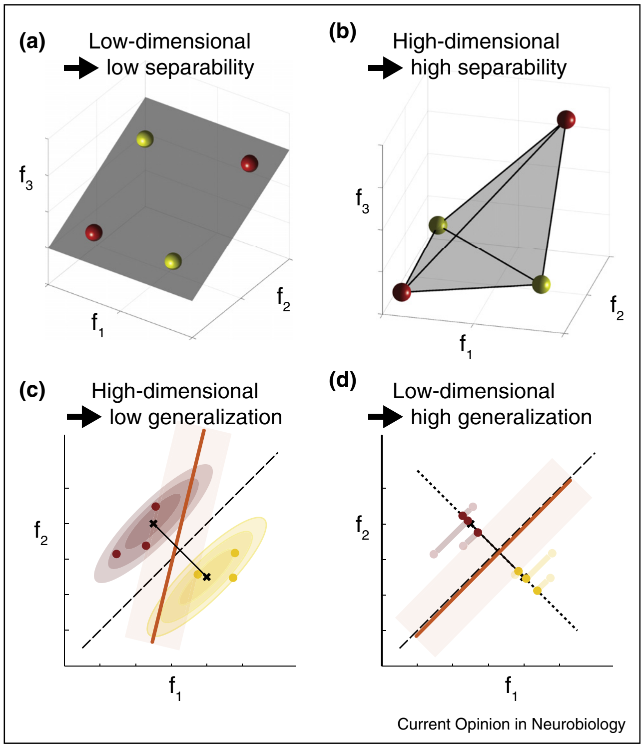
The traditional view of brain function is that individual neurons and even whole brain areas are akin to gears in a clock. Each is thought to be highly specialized for specific functions. This, however, does not fit with many observations, especially in higher-order cortex. For example, training monkeys 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 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.

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 nonlinear 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.