Comment

Meaning from sensory information

Rodrigo Quian Quiroga [a,b,*]

[a] Centre for Systems Neuroscience, University of Leicester, UK
[b] Leibniz Institute for Neurobiology, Magdeburg, Germany

Our brain is perhaps the most complex and fascinating system in the known universe. Billions of neurons talk to each other, share their messages and allow us to perform incredible feats with such facility that it seems even trivial. We hardly notice that we can effortlessly do something as complex as seeing, hearing, moving, thinking, remembering, feeling or even being aware of our own selves. Interestingly, or perhaps paradoxically, the neurons that form our brains can interpret and make sense of each others' messages, but we, the system as a whole, the masters of this entire process, hardly understand how.

Since more than half a century we are able to listen to the firing of single neurons. Neuroscientists have developed methodologies to study the involvement of neurons in different areas in a large variety of tasks. But complex brain processes are clearly related to the activity of large neural populations and seeing one or a few neurons at a time does not tell much. Since more recently, though, we can record and isolate the activity of a few hundred, perhaps up to thousand neurons, but then a new challenge arises: we have to make sense of all these data. Tracking the firing of one or a few neurons while an animal perceives a well-controlled stimulus is relatively easy, but how should we proceed to understand how the firing of a large number of neurons determines that we are seeing a Bengal tiger and not a kitten, a jaguar or the face of our mother? To make sense of this data, neuroscientists have resorted to tools from other disciplines, particularly from physics and mathematics – or, in more than a few cases, physicists and mathematicians who mastered these tools have turned their eyes into this problem, eventually becoming neuroscientists.

One very powerful tool to study large neural populations is given by information theory and analogous developments in machine learning, with the so-called decoding algorithms [9]. The basic idea of these methods is to estimate the amount of information about certain stimuli, in the case of information theory, or predict the particular stimuli, in the case of decoding, from the activation pattern of a population of neurons. Both these approaches are intrinsically related and have been shown to provide much more information than when considering single neurons in isolation (for a review see [9]). A radically different way to obtain information about whole populations is to directly record signals reflecting the population activity, as it is the case with electroencephalograms (EEGs) – recorded non-invasively from the scalp or intracranially – which represents the mean activity of relatively large population of neurons. Depending on specific details of the recording, even with a single EEG channel we have access to the common activity of thousands and up to millions of neurons, but now the problem is that: (i) we have lost the specificity of the activation patterns of single neurons, and (ii) the signal is much more noisy and challenging than in the case of single cell recordings.

Since the times of the first recordings by Hans Berger in the 1920's, neuroscientists are looking for different means to extract "hidden" information from EEG recordings [2]. Among these methods, stands out the study of dynamically evolving spatial patterns advanced by Walter Freeman and colleagues, which leads to the development of a neurodynamic formalism analogous to the Carnot cycle in thermodynamics [1]. Details of this formalism have been described elsewhere [1,2], so I will limit myself to discuss further the general issue addressed with this formalism of how the brain construct knowledge from perceptual information.

Our brain is constantly bombarded with a myriad of sensory inputs that trigger a cascade of activations leading to the attribution of meaning. We do not see wavelengths of different frequencies at different locations, as we don't hear sequences of time–frequency patterns or perceive minute pressure arrangements in our fingers. Instead, we see a cat, hear a friend's voice or touch an apple. This process of meaning attribution is very creative and subjective, far from being completely determined by the external stimulus per se. It is, in fact, constructed in our brains from relatively very poor information. We process only a minute amount of all the information received through our senses and we construct rich representations based on unconscious assumptions [3]. This attribution of meaning involves abstractions, extracting relevant features and leaving aside a countless number of details [6,7]. Let us illustrate this with the visual system. Neurons in primary visual cortex (V1) encode details of the stimulus, such as local orientations at particular locations [5]. Following the so-called *ventral visual pathway*, neurons in higher visual areas, particularly in the inferotemporal (IT) cortex, fire to a more complex representation of what is presented to the eye. A neuron in IT will not fire specifically to a given pixel intensity or the orientation of some line, but to a face, the shape of a hand, etc. [5,12]. Neurons in IT cortex do not care much about small differences in details and show some degree of visual invariance: if we slightly vary the picture, their firing will remain more or less the same [5,12]. Neurons in IT project to the hippocampus and surrounding cortex [4], what it is known as the medial temporal lobe. It is there that this process of abstraction – of encoding high level concepts and leaving aside minor details – reaches its pinnacle in the so-called concept cells [7,11]. These neurons represent particular concepts, *die dinge an sich*, and fire to different pictures of given persons (places, objects, etc.) and even to the person's written and spoken name [7,8].

The representation of meaning is implicit in primary sensory areas: a neuron in V1 cannot tell whether we are seeing a tiger, a chair or a friend's face; for this we need to consider the activity of many more neurons and put together the information of details that each of these neurons encode. In contrast, neurons in high level areas, like concept cells in the medial temporal lobe, give an explicit representation of the concept, as a single neuron can tell that we are seeing Jennifer Aniston and not a tiger or another actress [10]. This simple principle seems common to all sensory modalities. In general, sensory processing goes from a representation of details, an encoding of the physical attributes of the stimuli – and, in turn, an implicit representation of the concept – to an explicit representation of meaning in higher areas. Such explicit high level representation of concepts may indeed constitute a key neural substrate of human (and perhaps also animal) thought.

References

[1] Capolupo A, Freeman WJ, Vitiello G. Dissipation of 'dark energy' by cortex in knowledge retrieval. Phys Life Rev 2013;10:85–94 [in this issue].
[2] Freeman WJ, Quian Quiroga R. Imaging brain function with EEG: advanced temporal and spatial analysis of electroencephalographic signals. New York: Springer; 2012.
[3] Helmholtz H. The facts of perception. In: Kahl R, editor. Selected writings of Hermann von Helmholtz. Middletown (CT): Wesleyan University Press; 1878.
[4] Lavenex P, Amaral DG. Hippocampal–neocortical interaction: a hierarchy of associativity. Hippocampus 2000;10:420–30.
[5] Logothetis NK, Sheinberg DL. Visual object recognition. Annu Rev Neurosci 1996;19:577–621.
[6] Quian Quiroga R. Borges and memory: encounters with the human brain. Cambridge (MA): MIT Press; 2012.
[7] Quian Quiroga R. Concept cells: the building blocks of declarative memory functions. Nat Rev Neurosci 2012;13:587–97.
[8] Quian Quiroga R, Kraskov A, Koch C, Fried I. Explicit encoding of multimodal percepts by single neurons in the human brain. Current Biol 2009;19:1308–13.
[9] Quian Quiroga R, Panzeri S. Extracting information from neural populations: information theory and decoding approaches. Nat Rev Neurosci 2009;10:173–85.
[10] Quian Quiroga R, Reddy L, Koch C, Fried I. Decoding visual inputs from multiple neurons in the human temporal lobe. J Neurophysiol 2007;98:1997–2007.
[11] Quian Quiroga R, Reddy L, Kreiman G, Koch C, Fried I. Invariant visual representation by single neurons in the human brain. Nature 2005;435:1102–7.
[12] Tanaka K. Inferotemporal cortex and object vision. Annu Rev Neurosci 1996;19:109–39.