J Biosci (2022)47:46
DOI: 10.1007/s12038-022-00287-2

Clipboard

Face neurons and super-normal stimuli: A window into visual processing

Keywords. Artificial Intelligence; illusion; neural network; perception

How do we interpret patterns of light on our retina, and recreate from this our perception of landscapes, objects, and people? The classic way of thinking of this came from the pioneering work of Hubel and Wiesel (1959). Starting in the 1950s, they recorded neuronal activity from the visual cortex of cats and primates while the animals were given various visual stimuli. They presented these images by inserting slides into an old-fashioned slide projector. Quite by accident, they found that some cells responded best to the straight-line border of the slide as it was being loaded.

This led to a systematic characterization of responses, starting from simple orientation-selective cells (a line of light at a given angle), to cells that 'liked' the line when it went by at a particular speed, to angles, and so on. This very logical hierarchy of responses, each building in complexity by incorporating elements from simpler responses, became the dominant way to think about how the visual system reconstructs complex features. In this view, even complex stimuli like faces can be decomposed into a suitable combination of simpler features.

It should be pointed out that this idea did not go uncontested. One critique is the 'grandmother cell' argument (Barwich 2019). It points out that if cells indeed become more selective for stimuli as you go up the hierarchy, you should have a cell somewhere that responds to your grandmother. Not just that, but the cell would respond to your grandmother from a particular angle, smiling, side-lit, and so on. What would happen to your image of your grandmother if the cell died, or vice versa? In one case you could not recognize her. In the other, you would have a latent cell for a stimulus that would never come – indeed, you would have latent cells for all possible grandmothers. Clearly absurd.

The other note of caution came from an elegant neural network study by Lehky and Sejnowski (1988). They 'taught' a neural network a very simple task: to decide if an object in an image was convex or concave. This is a common illusion: we perceive bumps in an image as convex or concave depending on the orientation of the image with respect to the light. The network did this, but more interesting was what happened in the 'cells' of the network. Quite spontaneously, cells in the middle layers of the network took on just the same properties as Hubel and Weisel had seen: orientation selective, edge selective, and so on. Sejnowski argued that orientation selectivity was just an epiphenomenon, an accidental side-effect of quite a different computational goal. Despite this caution, the idea of high-order feature-selective cells is appealing, and there was quite a stir some years back when it was reported that certain primate and even human neurons respond to specific faces (Quiroga *et al.* 2005), and hence the term 'Jennifer Aniston neuron', which provides a cultural context for some of these studies.

Thus, we come to a recent report by Bardon *et al.* (2022) titled 'Face neurons encode nonsemantic features'. The core question in this study is whether the neurons encode faces (a semantic category) or simply respond to some correlation of visual features that happens to occur in the images of faces. If the former, then the cells should respond only to stimuli that are faces. Exceptions to this 'semantic' view of face neurons have already been reported to some extent by showing that these neurons respond somewhat to round objects and other subsets of

Published online: 14 July 2022

face features. It could be argued that the cell remains a face neuron, but since the input is a degraded image of a face, it is just a bit less active.

One way to unambiguously resolve this is to devise non-face stimuli that are as good, or even better, at causing the face neuron to respond. Super-normal stimuli are well known in the behavioral literature. For example, herring-gull chicks peck at a red dot on their parents' bills to get them to provide food. Tinbergen (1953) found that the chicks peck even more vigorously at a stick with an exaggerated red dot than at their parents' bills. If we can get neurons to respond in this exaggerated way to a non-face stimulus, then clearly it is a combination of features and not 'faceness' that triggers the cell.

How does one generate such a stimulus? An exhaustive search through the very high-dimensional space of possible stimuli is not feasible. Instead, Bardon *et al.* used a closed-loop Artificial Intelligence (AI) approach termed XDream (Ponce *et al*. 2019). They record from a neuron, use its responses to adapt the stimulus, and go through many cycles of this to iteratively come up with a super-stimulus. At the end of such a search, one has a set of images that triggers a very strong response in a face neuron.

Do these images look like faces? No, not to me! Rather than ask my subjective opinion, the authors check this point systematically by asking human subjects to classify these and other images in single words. In a series of tests, they find that the super-face stimuli do not seem to humans to be particularly face-like, no more so than dog images in one test. However, the images do look (to me) like how some abstract surreal artist might possibly draw a face. When asked for a one-word description of such images, humans classify them as faces about 4% of the time. Not high, but clearly above chance, since there are an enormous possible number of other words that could be used to label an image.

As in all such studies, there are some possible caveats. First, the XDream procedure was applied to neurons recorded from monkeys, but it was humans who categorized the resultant images. Second, perception is clearly not a single-neuron phenomenon, but rather one which involves multiple levels of networks in the brain. To their credit, the authors discuss these points in some detail.

Stepping back, it is worth noting that face recognition is important for survival in many organisms, especially social animals, which rely on visual information to recognize conspecifics, and in many cases even parse their emotional state. The 'Thatcher Illusion' (Thompson 1980) is a striking example of how there is clearly specialized circuitry to do this. It compares an inverted and upright facial image: both have the same geometrical elements, but we read much more into the upright one. I won't spoil the impact of the illusion with further explanation, but see *http://thatchereffect.com/*.

I will leave with one last thought experiment. We now have AI systems that do an excellent job of finding faces in photographs – these now even run in your phone camera. Face recognition programs are in the vanguard of AI research. Do these too have 'neurons' that respond to super-face stimuli? I would be curious to know what a super-face image for an AI looks like to a human.

References

Bardon A, Xiao W, Ponce CR, Livingstone MS and Kreiman G 2022 Face neurons encode nonsemantic features. *Proc. Natl. Acad. Sci. USA* **119** e2118705119

Barwich AS 2019 The value of failure in science: The story of grandmother cells in neuroscience. *Front. Neurosci.* **13** 1121

Hubel DH and Wiesel TN 1959 Receptive fields of single neurons in the cat's striate cortex. *J. Physiol.* **148** 574–591

Lehky SR and Sejnowski TJ 1988 Network model of shape-from-shading: neural function arises from both receptive and projective fields. *Nature* **333** 452–454

Ponce CR, Xiao W, Schade PF, *et al.* 2019 Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences. *Cell* **177** 999–1009

Quiroga RQ, Reddy L, Kreiman G, Koch C and Fried I 2005 Invariant visual representation by single neurons in the human brain. *Nature* **435** 1102–1107

Tinbergen N 1953 *The herring gull's world; a study of the social behaviour of birds* (London: Collins)

Thompson P 1980 Margaret Thatcher: a new illusion. *Perception* **9** 483–484

US Bhalla
National Centre for Biological Sciences,
Tata Institute of Fundamental Research,
Bengaluru 560065
India
(Email: bhalla@ncbs.res.in)