Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews 29 (2019) 111–114

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Comment

Akakhievitch revisited
Comment on "The unreasonable effectiveness of small neural ensembles in high-dimensional brain" by Alexander N. Gorban et al.

Rodrigo Quian Quiroga

Centre for Systems Neuroscience, University of Leicester, UK

Received 25 February 2019; accepted 27 February 2019
Available online 4 March 2019
Communicated by L. Perlovsky

Gorban, Makarov and Tyukin (GMT), three noted Russian scientists, are very well placed to revive the infamous story of Akakhi Akakhievitch, a neurosurgeon from the distant Ural Mountains, who, convinced that ideas are represented by specific neurons, performed experiments in animals (and in Trotskyists waiting for a death sentence) trying to localize the neurons representing the concept "mother". Once he found these neurons, he proceeded to ablate them in a person who was determined to forget his unbearable mother. The surgery was a success and when the person recovered, he had completely lost the notion of his mother; all memories related to her were gone. Reasoning that "mother" is an emotionally charged concept, Akakhievitch then decided to go for "grandmother cells".

The story, of course, is a hoax. It was conceived by neuroscientist Jerry Lettvin to animate a course he used to teach in the 1960s at MIT, to ridicule the idea that single neurons can encode specific concepts [1,2]. As GMT discuss in their article, the idea of neurons encoding specific things goes back to William James and his "pontifical cells"; something similar was discussed by Freud, when he talked about "psy and phi cells". Later on, the notion of pontifical cells was dismissed by Sherrington, who argued that the brain rather uses a "millionfold democracy", and it was subsequently reinstated by Jerzy Konorki, who argued for the existence of "gnostic cells", as well as by Horace Barlow, who supported the idea of "cardinal cells", i.e. small assemblies of neurons encoding percepts (for a short review on this discussion, see [3]). But, by far, what really caught the attention of neuroscientists was Lettvin's story of grandmother cells, which triggered very hot debates in the field [4–6]. Oddly enough, the only written record that exists from Lettvin's thoughts on grandmother cells is a personal letter he wrote to Horace Barlow, published as an appendix in a book chapter [7] (also reproduced in [1]).

Grandmother cells in the human MTL?

Patients suffering from epilepsy refractory to medication may be implanted with intracranial electrodes in the medial temporal lobe (MTL; the hippocampus and surrounding cortex) to evaluate the possibility of a surgical resection of the epileptic focus. These patients stay in hospital for about a week, and, through microwires protruding a couple millimetres from the intracranial electrodes, it is possible to record the activity of dozens of neurons while they perform different tasks [8]. With this setup we found neurons with a remarkable degree of selectivity and invariance – for

DOI of original article: https://doi.org/10.1016/j.plrev.2018.09.005.
E-mail address: rqqg1@le.ac.uk.

https://doi.org/10.1016/j.plrev.2019.02.014

example, one neuron fired to the presentation of 7 different pictures of Jennifer Aniston but not to 80 pictures of other persons, places or objects [9]. Moreover, these neurons also fired to the spoken and written name of the particular person, but not to the pictures or the names of other persons [10,11]. This evidence shows that such "Concept Cells" (a.k.a. "Jennifer Aniston neurons") respond to specific concepts, and not to the sensory features of the stimuli, and they might be the long-awaited grandmother cells. But let us analyse the evidence in more detail and also revise what we mean by "grandmother cell". After all, many of the controversies about grandmother cells arise from using the same term with different meanings.

The most naïve version of grandmother cells is that one (and only one) neuron encodes each concept: one for Jennifer Aniston, another one for Luke Skywalker and yet another one for grandma. But this cannot be the case of Concept Cells, as the chance to hit the only one neuron encoding Jennifer Aniston among billions others is practically zero. A more reasonable possibility is that many neurons (in his story, Lettvin postulated 18,000) may respond to one and only one concept. This is, however, impossible to prove as we cannot show all concepts in an experiment. So, if we find a neuron that responds only to Jennifer Aniston, we cannot rule out that this neuron may also respond to other concepts we did not show. In fact, the same neuron that was activated by pictures Jennifer Aniston, the next day also responded to Lisa Kudrow, a co-star of the TV series "Friends". This was not a "Friends neuron", though, because it did not respond to other actors of the same TV series. Perhaps we could retune our definition of "concept" and argue that the neuron encoded "the blonde actresses in Friends", but then we are dragged into an endless semantic discussion. The answer is clear, however, when we consider other evidence.

It is quite common to find that Concept Cells respond to more than one concept [11] and, when this is the case, these concepts tend to be associated [12]. So, it seems more reasonable to posit that Concept Cells encode associated concepts instead of searching for ad-hoc categorizations that may encompass the neuron's responses as a single, though relatively convoluted, concept. Moreover, in a recent study we showed that Concept Cells change their tuning to encode new associations between initially unrelated concepts – for example, a neuron initially firing to a certain person very rapidly started also firing to a particular place, once the association between the person and the place was arbitrarily created [13]. It could also be argued that Concept Cells may encode a single concept and respond (to a lesser degree) to other ones due to similarity with the first one [4]. However, it is very difficult to define similarity in such high-level cognitive space [14] and, furthermore, a recent study showed that when Concept Cells respond to more than concept, these responses are in most cases indistinguishable from each other [15].

The problem when engaging in the discussion of whether Concept Cells are grandmother cells or not is that we tend to overlook the crucial point: what is the function of these neurons? So far, all the evidence shows that these neurons are involved in episodic memory, encoding associations to, for example, remember meeting a certain person in a certain place [11]. They do so by expanding their tuning – i.e. a neuron initially firing to a given concept will start firing to another one that has been associated with the first [13]. The sole notion of grandmother cells excludes this possibility, because grandmother cells cannot change their tuning by definition.

These clarifications are also relevant when considering modelling works. A large number of studies have consistently shown that the MTL is involved in declarative memory and not in perception [16]. In line, Concept Cells can hardly be involved in perception, given that they do not carry information about the features of the stimulus – they fire to completely different pictures of a person and even to his/her written or spoken name; moreover, based on the neuron's firing we can decode that a particular concept was shown, but not which picture of this concept [17]. Therefore, models of how neurons may achieve robust discrimination of concepts seem more suitable to explain perceptual functions of the neocortex, which in turn send the results of its computations to the MTL. Furthermore, modelling studies demostrate that the rapid formation of associations showed by Concept Cells relies on very sparse representations (but not grandmother cell coding) [18,19], which is exactly what we find in our recordings.

Akakhievitch revisited

Let us finally illustrate the function of Concept Cells by giving a little twist to Lettvin's story. Drogori Akakhievitch, a young neuroscientist, grandson of the great neurosurgeon ridiculed by Lettvin, is determined to clear the name of his grandfather. He resolves to repeat his grandfather's major feat, ablating every single neuron representing a specific idea. But in spite of major technical advances in recent decades, this is far from an easy task, with his grandfather long dead and his experiments known only through the story told by Lettvin. In particular, he doesn't know where to look for these neurons, but the recent finding of Concept Cells in the human hippocampus gives him a good starting

point. He comes across a tormented patient who is desperate to forget, in this case, his ex-wife. Drogori obliges and proceeds to identify and ablate every single Concept Cell representing her, and this is more or less the conversation that takes place after the patient recovers from the surgery:

– Do you know who I am?
– Dr. Akakhievitch.
– Do you know what year it is?
– 2019.
– Now, think carefully... Are you married?
– Yes, well... not anymore...
– Do you remember your ex-wife?
– Yes.
(Silence... Akakhievitch is surprised; he shows him a picture of her).
– Do you recognize her?
– Yes.
(Long silence... This is not what he expected; not like the description of grandmother cells by his grandfather)
– Tell me about her.
– She is my ex-wife.
– And you are sure you know her.
– Yes, of course.
– Do you remember when the picture was taken?
– No.
– But it must have been you the one who took the picture. (Akakhievitch is confused)
– Which picture?
– The one of your ex-wife I just showed you.
– You didn't show me a picture of my ex-wife.
– What did I ask you a few minutes ago?
– If I knew you, which year it is, if I'm married...
– And I asked you about your ex-wife.
– No, you didn't.

Along these lines followed the conversation; the patient remembered facts (semantic memories) related to his ex-wife, he could recognize her (these being neocortical, not hippocampal functions), but he could not form or retrieve any episodic memory involving her. Unlike his grandfather, Drogori Akakhievitch could not erase the memory of the patient's ex-wife but he deleted all personal experiences (episodic memories) related to her. In a bizarre way, he may have given him a chance of reconciliation.

Conclusions

This comment tried to clarify that Concept Cells do show an extremely sparse (and invariant) coding, but should not be mistaken for grandmother cells. It should also be stressed that Concept Cells are not involved in identifying a particular stimulus or concept. They are rather involved in creating and retrieving associations and can be seen as the "building blocks of episodic memory" [11]. Most importantly, I commend GMT's effort to model relatively recent findings with Concept Cells. Such studies and further interactions between experimenters and modellers may clarify the function of these neurons, which so far have not been identified in other species and could be the basis of our unique human thoughts and intelligence.

References

[1] Gross C. Genealogy of the "grandmother cell". Neuroscientist 2002;8:512–8.
[2] Quian Quiroga R, Fried I, Koch C. Brain cells for grandmother. Sci Am 2013;308(2):30–5.
[3] Quian Quiroga R. Gnostic cells in the 21st century. Acta Neurobiol Exp 2013;73:1–9.
[4] Bowers J. On the biological plausibility of grandmother cells: implications for neural network theories in psychology and neuroscience. Psychol Rev 2009;116:220–51.
[5] Page M. Connectionist modelling in psychology: a localist manifesto. Behav Brain Sci 2000;23:443–512.

[6] Rose D. Some reflections on (or by?) grandmother cells. Perception 1996;25:881–6.
[7] Barlow HB. The neuron doctrine in perception. In: Gazzaniga M, editor. The cognitive neurosciences. Boston: MIT Press; 1994.
[8] Rey H, et al. Single cell recordings in the human medial temporal lobe. J Anat 2015;227:394–408.
[9] Quian Quiroga R, et al. Invariant visual representation by single neurons in the human brain. Nature 2005;435(7045):1102–7.
[10] Quian Quiroga R, et al. Explicit encoding of multimodal percepts by single neurons in the human brain. Curr Biol 2009;19:1308–13.
[11] Quian Quiroga R. Concept cells: the building blocks of declarative memory functions. Nat Rev Neurosci 2012;13:587–97.
[12] De Falco E, et al. Long-term coding of personal and universal associations underlying the memory web in the human brain. Nat Commun 2016;7:13408.
[13] Ison M, Quian Quiroga R, Fried I. Rapid encoding of new memories by individual neurons in the human brain. Neuron 2015;87:220–30.
[14] Quian Quiroga R, Kreiman G. Measuring sparseness in the brain: comment on Bowers (2009). Psychol Rev 2010;117:291–9.
[15] Rey H, et al. Encoding of long-term associations through neural unitization in the human medial temporal lobe. Nat Commun 2018;9:4372.
[16] Squire L, Zola-Morgan S. The medial temporal lobe memory system. Science 1991;253:1380–6.
[17] Quian Quiroga R, et al. Decoding visual inputs from multiple neurons in the human temporal lobe. J Neurophysiol 2007;98(4):1997–2007.
[18] Marr D. Simple memory: a theory for archicortex. Proc R Soc Lond, B 1971:23–81.
[19] McClelland JL, McNaughton BL, O'Reilly RC. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol Rev 1995;102:419–57.