

- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115–147.
- Bowers, J. S. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review*, 116, 220–251.
- Connor, C. E., Brincat, S. L., & Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Current Opinion in Neurobiology*, 17, 140–147.
- DiCarlo, J. J., & Cox, D. (2007). Understanding invariant object recognition. *Trends in Cognitive Sciences*, 11, 333–341.
- Gross, C. G. (2008). Single neuron studies of inferior temporal cortex. *Neuropsychologia*, 46, 841–852.
- Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Physiology (London)*, 35, 96–111.
- Hahnloser, R. H. R., Kozhevnikov, A. A., & Fee, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, 419, 65–70.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational properties. *Proceedings of the National Academy of Sciences, 7 USA*, 9, 2554–2558.
- Hopfield, J. J. (2007). Hopfield network. *Scholarpedia*, 2, 1977.
- Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005, November). Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310, 863–866.
- Koch, C. (2004). *The quest for consciousness*. Englewood, NJ: Roberts.
- Kreiman, G., Hung, C. P., Kraskov, A., Quian Quiroga, R., Poggio, T., & DiCarlo, J. J. (2006). Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. *Neuron*, 49, 433–445.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: 1. An account of basic findings. *Psychological Review*, 88, 375–407.
- McClelland, J. L., Rumelhart, D. E., & Group, P. R. (1986). *Parallel distributed processing: Psychological and biological models* (Vol. 2). Cambridge, MA: MIT Press.
- Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, 335, 817–820.
- Mormann, F., Kornblith, S., Quian Quiroga, R., Kraskov, A., Cerf, M., Fried, I., & Koch, C. (2008). Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *Journal of Neuroscience*, 28, 8865–8872.
- Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, 110, 611–646.
- Olshausen, B. A., & Field, D. J. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14, 481–487.
- Perez-Orive, J., Mazor, O., Turner, G. C., Cassenaer, S., Wilson, R. I., & Laurent, G. (2002, July). Oscillations and sparsening of odor representations in the mushroom body. *Science*, 297, 359–365.
- Quian Quiroga, R. (2007). Spike sorting. *Scholarpedia*, 2, 3583.
- Quian Quiroga, R., Kreiman, G., Koch, C., & Fried, I. (2008). Sparse but not “grandmother-cell” coding in the medial temporal lobe. *Trends in Cognitive Sciences*, 12, 87–91.
- Quian Quiroga, R., Nadasdy, Z., & Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation*, 16, 1661–1687.
- Quian Quiroga, R., & Panzeri, S. (2009). Extracting information from neural populations: Information theory and decoding approaches. *Nature Reviews Neuroscience*, 10, 173–185.
- Quian Quiroga, R., Reddy, L., Koch, C., & Fried, I. (2007). Decoding visual inputs from multiple neurons in the human temporal lobe. *Journal of Neurophysiology*, 98, 1997–2007.
- Quian Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435, 1102–1107.
- Rieke, F., Warland, D., de Ruyter van Steveninck, R. R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.
- Rolls, E. T., Treves, A., & Tovee, M. J. (1997). The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. *Experimental Brain Research*, 114, 149–162.
- Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89, 60–94.
- Rumelhart, D. E., McClelland, J. L., & Group, P. R. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations*. Cambridge, MA: MIT Press.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, 19, 109–139.
- Waydo, S., Kraskov, A., Quian Quiroga, R., Fried, I., & Koch, C. (2006). Sparse representation in the human medial temporal lobe. *Journal of Neuroscience*, 26, 10232–10234.
- Willmore, B., & Tolhurst, D. J. (2001). Characterizing the sparseness of neural codes. *Network: Computation in Neural Systems*, 12, 255–270.
- Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N., & Suzuki, W. A. (2003, June). Single neurons in the monkey hippocampus and learning of new associations. *Science*, 300, 1578–1581.
- Young, M. P., & Yamane, S. (1992, May). Sparse population coding of faces in the inferior temporal cortex. *Science*, 256, 1327–1331.

Received February 9, 2009

Revision received June 8, 2009

Accepted June 9, 2009 ■

## Postscript: About Grandmother Cells and Jennifer Aniston Neurons

Rodrigo Quian Quiroga  
University of Leicester

Gabriel Kreiman  
Children's Hospital Boston, Harvard Medical School, and  
Harvard University

A typical problem in any discussion about grandmother cells is that there is not a general consensus about what should be called as such. Here, we discuss possible interpretations in turn and

contrast them with what we find in our own data (arguably the closest experimental evidence of grandmother cells so far). A first and naïve interpretation of the term grandmother cell is that one and only one neuron encodes for one and only one concept (a face, an object, an animal, etc.). We agree with Bowers (2010) that this is a straw-man version of this idea—although some people still take this view when (incorrectly) arguing that if we would have grandmother cells then the concept of grandma would disappear if her dedicated cell dies—which clearly does not apply to our data. Given that we record from a very tiny fraction of neurons in the medial temporal lobe (MTL), if we do find a neuron firing to a particular concept, there must be more. A more plausible version

of grandmother cells is to say that many neurons code for one and only one concept. A typical theoretical argument against this view is what is known as *combinatorial explosion*: There are not enough neurons to code for all possible concepts (grandma seen from the front, in profile, with a blue hat, etc.). But this criticism can be easily refuted if one considers (a) invariance, in the sense that neurons do not need to encode all possible instances of a given concept and can instead encode abstract representations (Quian Quiroga, Reddy, Kreiman, Koch, & Fried, 2005) and (b) personal relevance, in the sense that neurons may encode only those things that are relevant to the person (a florist may distinguish between different types of orchids, but for an average person these may be all the same concept). In this respect, we recently found that personally relevant persons are more likely to elicit responses in the human MTL (Viskontas, Quian Quiroga, & Fried, in press). But, although we do not rule out this second interpretation of grandmother cells as implausible, it is not what we find in our data either. We showed a neuron firing to two different basketball players and another one firing to two different landmarks (but not to other landmarks; Quian Quiroga & Kreiman, 2010). Adding to this, we showed several other such examples (Quian Quiroga, Kraskov, Koch, & Fried, 2009): a neuron firing to Luke Skywalker and Yoda, both characters of Star Wars; another three neurons firing to two or more researchers performing experiments with the patients; and another firing to a spider and a snake (but not to other animals), and so on.

Following the previous line of reasoning, one could still argue that since the pictures the neurons fired to are related, they could be considered the same concept, in a high level abstract space: “the basketball players,” “the landmarks,” “the Jedi of Star Wars,” and so on. But now the whole issue of whether these are grandmother cells or not is just semantic. For now, let us leave aside the heavy connotations of the term *grandmother cell* and focus on characterizing the properties and functions of what have been dubbed *Jennifer Aniston neurons*. First, these neurons show a very abstract representation, firing selectively to completely different pictures of the same person: for example, a neuron in the hippocampus fired to seven different pictures of Jennifer Aniston but not to 80 other pictures of different people, objects, or animals (Quian Quiroga et al., 2005). This level of abstraction goes beyond a specific sensory modality since these neurons can also selectively fire to the person’s written and spoken names (Quian Quiroga et al., 2009). Second, the representation by Jennifer Aniston neurons is extremely sparse, given that they fire to very few of the pictures shown (Quian Quiroga, Reddy, Koch, & Fried, 2007; Quian Quiroga et al., 2005). Third, the representation by these neurons is explicit, in the sense that from the activity of a few of them it is possible to tell which picture is shown well above chance (Quian Quiroga & Panzeri, 2009; Quian Quiroga et al., 2007). In terms of function, given the long latencies of these responses (Mormann et al., 2008; Quian Quiroga et al., 2005), previous knowledge from patient H.M., and related findings showing that the hippocampus is not involved in perception but rather in memory (Squire, Stark, & Clark, 2004), we have postulated that these neurons have an abstract encoding to convert perceptions into memories (Quian Quiroga, Kreiman, Koch, & Fried, 2008). If this is the case, then it is not surprising that they fire to concepts that are related, given that such encoding of associations is a basic mechanism for learning and the creation of new memories. This interpretation complements stud-

ies in monkeys showing the firing of MTL neurons to formed associations (Miyashita, 1988; Wirth et al., 2003).

How the brain processes sensory inputs to create full percepts that lead to motor outputs, memory formation, and behavior in general is a fascinating question in neuroscience. As mentioned in our comment, there is likely not only one type of representation across all levels of sensory processing, and evidence from other species points toward a convergence from distributed to sparse representations (see e.g., (Hahnloser, Kozhevnikov, & Fee, 2002; Perez-Orive et al., 2002). As pointed out by Barlow, Parker, Singer, and Thorpe, 2009), information about physical stimuli is naturally coded in a distributed manner because this is the way external physical signals impinge on the sensory receptors. In this respect, we observed that all visual information is present in the retina, though with a distributed and implicit code (i.e., from the firing of a single cell in the retina, we cannot typically tell which stimulus is present). Bowers (2010) claimed in his reply that a great deal of information is not present in the retina and that higher levels of the system do not simply rerepresent retinal information but add new information. However, this statement violates the data processing inequality, which basically states that no processing of given data—as the one done by neurons in higher sensory areas—can increase the amount of information (Cover & Thomas, 2001; Quian Quiroga & Panzeri, 2009). Higher sensory areas are not just adding information but are making the internal representation about the stimulus (i.e., our perception) explicit (see e.g., Quian Quiroga, Mukamel, Isham, Malach, & Fried, 2008). In this respect, we proposed to measure the degree of sparseness, to quantify where in the continuum between distributed and sparse coding given representations lay.

According to the above mentioned view, one may wonder how sparse is the representation in the MTL at the very end of the pathway processing visual information (Felleman & Van Essen, 1991). In other words, how many neurons respond to a given stimulus, and conversely, to how many stimuli does a neuron respond? For this, we used a Bayesian analysis to estimate that less than a few million neurons respond to a stimulus and that each of them could fire to up to a few dozen pictures (Waydo, Kraskov, Quian Quiroga, Fried, & Koch, 2006). But as we clearly stated (Quian Quiroga & Kreiman, 2010; Quian Quiroga, Kreiman, et al., 2008; Waydo et al., 2006), these numbers should be taken as upper thresholds because (a) we used familiar stimuli, given that these are more likely to elicit responses (Viskontas et al., in press), and (b) neurons with higher degrees of sparseness may have been missed in our relatively short recording sessions. We showed in our comment that Bowers’s (2009) two main arguments in his target article for considering our estimation flawed (different from the ones mentioned above) were incorrect. In his reply, Bowers (2010) argued that our premises are suspect because (a) many neurons may be involved in coding for the familiarity of an object, and it would be a mistake to average the sparseness of all neurons to reach a single estimate; (b) it is more difficult to identify neurons with higher degrees of sparseness, compared to identifying those with lower degrees of sparseness; (c) our a priori estimate of sparseness (before any data were collected) was that all sparseness values are equally likely, thus being biased against grandmother cells. Let us answer to these points in turn. First, the issue of familiarity was mentioned before, but concerning Bowers’s (2010) point that it is a mistake to average neurons potentially

encoding for familiarity, note that we did not only report the mean sparseness value (0.54), but also the peak of the distribution (0.23), which is not affected by these—relatively few—neurons with low selectivity. We also discussed the bias given by the fact that we may miss very selective neurons, but we have to base our estimations on the neurons we record. Third, considering all possible sparseness values equally likely does not introduce the bias mentioned by Bowers (2010). After all we could have found that all neurons have the maximum possible selectivity. But of course, the maximum selectivity value we could measure is bounded by the total number of pictures used (e.g., we cannot get less than 1% responses with 100 pictures). We also note that the estimate of sparseness given in Bowers's (2010) example with a million cells firing to only 1 stimulus is incorrect because it does not use the Bayesian formulation considering the number of recorded neurons.

Summarizing the previous point, the arguments in Waydo et al. (2006) were not given to rule out grandmother cells but rather to estimate an upper threshold to the size of a network encoding percepts and the number of stimuli each neuron fires to. The fact that more than a neuron responds to one concept and that neurons do not necessarily respond to only one concept are given by the data itself (Quian Quiroga, Kreiman, et al., 2008; Quian Quiroga et al., 2005). There is still one final issue to comment on. Contrary to Bowers (2010), we do not make much distinction between what a neuron codes for and what it responds to. The emphasis made by Bowers (2010) in whether a neuronal response is interpretable is what we measure with decoding algorithms or information theory (Quian Quiroga & Panzeri, 2009). But due to trial-by-trial variability, noise, and so on, these algorithms do not offer yes–no answers. For example, from the firing of our MTL neurons we could decode picture presentations way above chance, but in the example shown in Figure 3 of Quian Quiroga et al. (2007), only 1 out of 32 pictures eliciting responses in this session could be decoded perfectly. Performance for the other pictures was 66% or lower (still much better than chance), reflecting the fact that many neurons responded to more than 1 stimulus.

### References

- Barlow, H. B., Parker, A., Singer, W., & Thorpe, S. (2009). Barlow's 1972 paper. *Perception*, *38*, 795–807.
- Bowers, J. S. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review*, *116*, 220–251.
- Bowers, J. S. (2010). More on grandmother cells and the biological implausibility of PDP models of cognition: A reply to Plaut and McClelland (2010) and Quian Quiroga and Kreiman (2010). *Psychological Review*, *117*, 300–308.
- Cover, T. M., & Thomas, J. A. (2001). *Elements of information theory*. Hoboken, NJ: Wiley.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1–47.
- Hahnloser, R. H. R., Kozhevnikov, A. A., & Fee, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, *419*, 65–70.
- Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, *335*, 817–820.
- Mormann, F., Kornblith, S., Quian Quiroga, R., Kraskov, A., Cerf, M., Fried, I., & Koch, C. (2008). Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *Journal of Neuroscience*, *28*, 8865–8872.
- Perez-Orive, J., Mazor, O., Turner, G. C., Cassenaer, S., Wilson, R. I., & Laurent, G. (2002, July). Oscillations and sparsening of odor representations in the mushroom body. *Science*, *297*, 359–365.
- Quian Quiroga, R., Kraskov, A., Koch, C., & Fried, I. (2009). Explicit encoding of multimodal percepts by single neurons in the human brain. *Current Biology*, *19*, 1308–1313.
- Quian Quiroga, R., & Kreiman, G. (2010). Measuring sparseness in the brain: Comment on Bowers (2009). *Psychological Review*, *117*, 291–299.
- Quian Quiroga, R., Kreiman, G., Koch, C., & Fried, I. (2008). Sparse but not “grandmother-cell” coding in the medial temporal lobe. *Trends in Cognitive Sciences*, *12*, 87–91.
- Quian Quiroga, R., Mukamel, R., Isham, E. A., Malach, R., & Fried, I. (2008). Human single-neuron responses at the threshold of conscious recognition. *Proceedings of the National Academy of Sciences USA*, *1*, 3599–3604.
- Quian Quiroga, R., & Panzeri, S. (2009). Extracting information from neural populations: Information theory and decoding approaches. *Nature Reviews Neuroscience*, *10*, 173–185.
- Quian Quiroga, R., Reddy, L., Koch, C., & Fried, I. (2007). Decoding visual inputs from multiple neurons in the human temporal lobe. *Journal of Neurophysiology*, *98*, 1997–2007.
- Quian Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*, 1102–1107.
- Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, *27*, 279–306.
- Viskontas, I., Quian Quiroga, R., & Fried, I. (in press). Human medial temporal lobe neurons respond preferentially to personally relevant images. *Proceedings of the National Academy of Sciences of the United States of America*.
- Waydo, S., Kraskov, A., Quian Quiroga, R., Fried, I., & Koch, C. (2006). Sparse representation in the human medial temporal lobe. *Journal of Neuroscience*, *26*, 10232–10234.
- Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N., & Suzuki, W. A. (2003, June). Single neurons in the monkey hippocampus and learning of new associations. *Science*, *300*, 1578–1581.