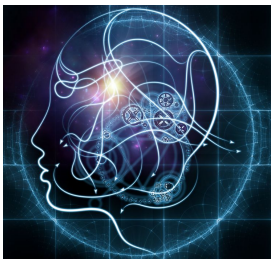

CS/NEUR125 Brains, Minds, and Machines

Assignment 1: Grandmother cells

Due: Thursday, February 23 by 9:00pm

This Assignment is a guided reading of the 2013 Scientific American article, "Brain Cells for Grandmother," together with a 2005 primary research article by Quiroga and colleagues entitled, "Invariant visual representation by single neurons in the human brain." Reading these articles will prepare us to discuss them during our first "Journal Club" during class on Friday, February 24.

To begin, create a copy of this Google document, modify the title of the copy to include your name, as you've done for labs (for review, see the Working with Google Docs for Lab Handouts web page--and note that you'll work individually rather than in pairs). Questions that you should submit answers to are shown in blue. As with labs, you'll turn in this Assignment by sharing your copy of the document with Ellen and Mike.

We recommend reading the Scientific American article first, as it is easier to read and gives a broader context for the experimental results that are reported in the 2005 paper. Reading a primary research paper is more challenging and takes some practice, so don't be discouraged if you're not understanding every line. Our goal is not to understand every line, but to explore the idea of "grandmother cells" and sparse representations, understand the main experimental results, and begin to appreciate some of the technical details.

Many of the questions below are trying to get you to look up information in specific paragraphs or sentences of the paper. If you use phrases from the paper in answering the questions, you must put them in quotation marks, and you should try to reformulate the idea in your own words. (If the answer is simply a name or a number this is not necessary.)

Over the course of the semester, we will use these non-lab assignments to develop your ability to express scientific ideas clearly and precisely. However, for this assignment our emphasis is on helping you understand the papers rather than testing your ability to interpret them on your own. That is to say we will grade generously so you should not be anxious about answering these questions perfectly.

I. Brain Cells for Grandmother

Q1. Briefly describe the grandmother cell hypothesis. What are the "extreme" and "less extreme" forms of it?

Q2. What did the authors discover about the "Jennifer Aniston neuron" the day after they first found it? How do they argue the cell can still be considered a "grandmother cell"?

Q3. About how many neurons do the authors say reside in the medial temporal lobe? About how many concepts do they say a typical person remembers? Do they cite a source for this latter estimate?

"Sparse firing" refers to a situation in which a very small fraction of neurons in a population are activated at any given time. Similarly, **sparse coding** or **sparse representations** refer to situations where the brain uses relatively few out of a large population of neurons to represent particular individuals, objects, or memories.

Q4. What do the authors cite as a possible useful function or advantage of sparse neural codes for important concepts or people? What is a potential disadvantage of using a broadly distributed code "with neurons coding for each minute detail," instead of a sparse code?

II. Invariant visual representation by single neurons in the human brain

In a primary research paper as compared to an article in the popular press, the authors are more likely to assume knowledge in the reader, or leave out details that are familiar to experts. We'll try to fill in some of the assumed knowledge with this document, but we'll also have to accept that we won't digest and understand every line in this technical paper.

So part of what we're practicing, is choosing where to put our reading effort in order to get what we want to get out of the paper. For now what we want is a basic understanding of their experimental results, and we want to learn enough about their methods to understand what they're trying to show in each figure. We will not need to look into the Supplementary Information for this, so don't bother looking that up online.

Because it's easy to get bogged down in technical details in a paper, you first want to understand what is the question or hypothesis the authors are trying address with their study. That way you can try to relate everything else you read to answering that question--and if it doesn't help address the main question, you might be able to safely ignore it (until you realize that it's relevant or important somehow).

Q5. What is the main question addressed by this paper?

Start with the title and abstract, and try to note unfamiliar terms. Sometimes they will be defined on a later page, sometimes they won't.

Q6. When the authors refer to an "invariant visual representation," what is it that is "invariant," or relatively unchanging?

Q7. What are some of the "metric properties" of an image that the Abstract says don't affect an "invariant visual representation"?

Q8. What task were subjects performing while viewing images? Why?

Q9. From the first paragraph after the Abstract, what does a "unit" refer to? Based on this, what do you think a "single unit" or "multi-unit" refer to?

The Abstract concludes that neurons in the medial temporal lobe (MTL) instantiate an "invariant, sparse, and **explicit** code." They are using the word "explicit" in a technical sense here, which they don't define until the last paragraph of the paper, just before the Methods.

Q10. What parts of the brain did the neurons recorded in this study come from?

Q11. How many units (of either kind) did the authors record from? What percentage of those responded significantly to at least one of their pictures? What were their criteria for a significant response? How many pictures did they present to each unit on average? (Hint: see the Methods section at the end of the paper.)

Q12. Figure 1A. What is plotted on the x and y axes? What is the main point of this figure?

Q13. Figure 1B. What is the "median response"? That is, what does "median" mean? From this figure, how many of the images that were not considered to be "pictures of Jennifer Aniston" appear to trigger a significant response from this neuron?

The authors say they used the Receiver Operating Characteristic (ROC) framework to quantify the "degree of invariance" in the responses of cells they thought responded to specific individuals, objects, or landmarks. By "degree of invariance" they mean how reliably this cell continues to respond despite changes in the way the individual is presented. In other words, does the cell respond to all instances of the individual or object, including images of them or their names?

More generally this ROC framework ***quantifies the quality of a binary classification based on a neuron's firing rate***. That is, with this framework we can ask, how accurately can we decide if a given stimulus *is or is not* an "X"--that's the "binary classification"--given a particular neuron's firing rate in response to the given stimulus? We assume that the decision is made by comparing our neuron's firing rate to a threshold; if higher than the threshold we say it is an X, if lower we say it is NOT an X. We have freedom to set the threshold for X-detection as high or low as we want though.

Now, our hit rate is the fraction of the time we say it is an X, when it really is. And our false positive rate is the fraction of times where we say it is an X when it really is NOT. The ROC curve (like those plotted in Figs 1C, 2C, 3C) shows the hit rate vs. false positive rate for every possible choice of threshold. If we set the threshold for X-detection high (left side of ROC curve) then we won't have many false positives--but we'll also miss some real Xs: our hit rate will suffer. If we set the threshold for X-detection low we'll catch more of the Xs, but we'll also catch more non-Xs: our false positive rate will get worse.

Since the number of correctly classified stimuli depends on our choice of threshold, "percent correct" is not the best way to quantify how well our neuron classifies stimuli. Instead, ***the area under the ROC curve (AUC) is a commonly used measure of classification quality. This AUC will be a number between 0.5 and 1 that corresponds to the fraction of correctly classified stimuli averaged over all possible choices of threshold. A perfect classifier neuron would have an AUC of 1.***

Figure 1C. The red ROC "curve" (i.e. rectangular corner) is a way of quantifying this neuron's ability to discriminate images of Jennifer Aniston from non-Aniston images. The area under the red curve is 1, indicating 100% correct classification and implying that this cell responded significantly to every picture of Jennifer Aniston and did not respond significantly to any picture that was not in the Jennifer Aniston "category" of images.

Q14. Figure 1C. The grey ROC curves quantify this neuron's ability to classify other categories of images. How were these other image categories defined by the authors? (Refer to the text as well as the figure caption.)

Q15. Figure 1C. Based on their "ROC surrogate curves" analysis, the authors report that "$P < 0.01$." P here stands for a probability. What is it the probability of? (Again, refer to the text to answer this question.)

Q16. Figure 2. What is the significance of the fact that this "Halle Berry unit" responds to photos or line drawings of Halle Berry but not to other line drawings of people?

Q17. What percentage of responsive units showed "invariant visual responses" to particular persons, animals, places or objects? What percentage of all the units they recorded is that?

Q18. Why do the authors say the invariant neuronal responses they observed "cannot be attributed to any particular movement artefact"? What do you think the authors mean by "movement artefact"?

Q19. Please submit two questions you have about terms, figures, concepts or anything in these two articles that confused you or that you'd like to pursue further during our Journal Club discussion. For example, one question might be related to a technical detail, and another might be broader (e.g. related to assumptions, methods, interpretation, or open questions for future research).