Hi guys,

I read your paper, and I don’t really understand how this overleaf thing works, and also read it on paper (as in, I printed it), so you will have to parse my comments from this email. It seems like a very cool paper!

General Comments

- I feel like the evidence favoring primacy coding is a big deal, but as far as I can see, is not highlighted in the abstract. Perhaps put it more front and center?

- line 222: I don’t think we assumed lower AND upper bounds on Epsilon in our elife paper. We only had one bound so that there was a maximum possible gain.

**Figure 1:**

I think this is a good introductory figure, though it verges on being a little too crowded. Maybe kill some panels (like the lower ones in D). DONE Also consider labelling the output of the bacterial model (like you did the output of the LN model). DONE I don’t think you ever defined the “complexity” of an odor, or, at the least, you used that term before you defined it. DONE Finally, I know what you’re trying to do with those tuning curves in E, but it’s pretty obscure, and most of the people reading your paper will not be familiar with that paper from John’s lab. So you have to explain that you reordered the odorants, etc. Mentioned in caption Also you have nothing on your Y axis in those panels. Similarly, X axis in D is not labelled. DONE

**figure 2:**

I struggled a lot with this one. Part of it, I think was because there’s a lot going on here. Moved MI figures to Supp Info.

a) You have to say in the legend that tau is the adaptation time in the bacterial chemotaxis model that you use as your front end.

b) I do not understand why MI decreases with concentration in some panels, but does not in some panels. There is no correlation between whether it decreases or not with whether you use the adaptive or non-adaptive model. If you have an intuition that is illuminating, you should explain this in the text.

c) Also, I have no idea how the panels in B and C are being made — specifically, where is the X-axis coming from? Are you varying the odor intensity in different “trials”? In all possible combinations? For both A and B? Not clear.

d) Visually, I can’t see the thin or thick lines in any of the panels. I just see a smear. So if it’s important to convey that, you have to emphasize that more. Similarly, I can’t see the small and big dots in the t-SNE embedding (though that is very convincing and is a nice figure!) Fixed

e) Why does the MI not peak at at the max. Sensitivity  in the adaptive case? (In fact, if you look carefully at the non-adaptive case, it doesn’t peak there either. You say in the text that MI peaks at 100. This is not the case in any of the panels in B or C. x-labels were labeled incorrectly

**Figure 3:**

a) You have ticks indicating odor signal above 5V in panel D. But you went on in the introduction about how odor intensities in the real world can be distributed over many intensities! In fact, in our eLife paper, we point out that the ORN can respond even to tiny whiffs. What you’re doing with this 5V cutoff is you’re only looking at the extreme biggest whiffs. If you care only about coding for these, then what’s the point of taking the naturalistic stimulus? And why do all this adaptation? Figure was cut off on the vertical scale to magnify the small whiffs visually, so there are actually a larger range of whiffs than is shown. This has been amended. The small whiffs are indeed quite small compared to some other larger ones. Smaller than 5V are just like 1 frame blips, so this is why chose 5V.

b) I had a very hard time understanding panel B. It has too many sub panels, and too many complicated axes labels. Agree. Reduced 18 panels to 4.

c) I don’t understand why there are no corners that are yellow in panel B. For example, when you have maximum foreground intensity and minimum foreground complexity, why can’t you decode all the odors? Shouldn’t that corner be at 100%? High foreground can still saturate the system, so decoding ability compromised

D) You have to explain what the trace in panel D is. It is mysterious to a lay reader.

E) Panel E is missing Y axis labels.

**Figure 4:**

I liked this result the most, and think this is really exciting.

a) There are no Y ticks or numbers in panel A

b) I would like to see another panel showing size of primacy set required to decode odor at some quality vs. odor complexity. Is this scaling sublinear? If so, that would be a very interesting finding. This could potentially replace panel C, which is very obscure.

c) In panel B, if odor step intensity is being reported as a fold change, (X), you don’t have to have arb. units.

d) I have no idea what the top thing in panel C is supposed to show. What is the x axis? (Or y axis?)

e) Can we increase the mesh resolution in panel B 100X? It would make for a much nicer picture.

**Figure 5:**

I understand this is the “big picture” result, as this is what matters to the brain, etc. but I am not convinced by this figure or the result. You set up this model (in a pretty parsimonious manner), and plot some traces of accuracy vs. # of odors and compare adaptation to no-adaptation. And the result is that adaptation (or divisive norm.) helps. That adaptation helps cf. no adaptation is not at all surprising, so the real novel result is here that ORN adaptation > divisive normalization. (Is that fair?) But I suspect that there are lots of free parameters in this very large model, and while you say that they were constrained by experiments, there’s still a lot of wiggle room, and I wonder how panel A would look if those knobs were changed a bit. I suspect that you can get these curves to be anywhere you want relative to each other without choosing ridiculous parameters in your model. So I am less convinced that this is generally true and more inclined to think that this is what you saw for some very particular (possible arbitrary) point in the parameter space of your huge model. I believe the result is valid. The network connection topology was chosen randomly (with statistics given by experiment). There are no other free parameters than this connection topology. Still, as you say, the toplogy may drive the performance. However, the classifier (connection weights from MB to readout layer) was trained on half of the data set. Since the learning is supervised, it should optimize the classifier connection weights, removing sensitivity to the particular instantiation of AL-MB topology.

Nonetheless, to verify that the AL-MB connection topology does not drive results, the results shown were repeated on 10 different AL-MB connection topologies, randomly chosen -- the plots show the average among these 10.

a) Perhaps this is my own ignorance, but don’t the detailed of the classifier you use matter? If your R vector is scrambled in some nonlinear way, a linear classifier would presumably do a bad job decoding it — but that doesn’t mean that R vector is garbage — right? Used a softmax cross-entropy which is the most straightforward way of doing multi-class classification in a principled way.

b) line 486: “odors of the same identity but different intensity were assigned the same valence”. This is not true in reality. Lots of odors are attractive at one concentration and repellant at a different concentration. True, will mention it’s an assumption Maybe you want to say this was a simplifying assumption?

**Minor suggestions**

- the clause separated by the m-dash in the abstract (“ORN gain scales inversely”…) is very hard to parse. Consider splitting into two sentences.

- line 90: “a simpleR system such as E coli…”

- line 74: I don’t believe in “super-sustained responses” (it’s not clear to me that it’s not a consequence of a very sticky odor) — do you really have to mention it?

- line 27: “contribute also significantly” —> “also contributeS significantly”

- Figure 1 caption: when you say “step”, you mean “pulse”

- line 181: don’t have a variable called A\_a. Very confusing.

- Figure 1D. You say the bottom panels are normalized, but the traces don’t reach 1. Also, what’s the point of showing normalized traces? If they show something that the non-normalized traces don’t, then say so.

- Figure 2D: consider dropping the axes on the t-SNE plots. They don’t mean anything.

I didn’t get to the supplemental info, but I hope this helps! Good luck with the submission, I think this is a very interesting paper!

Cheers,

Srinivas