Dear Editor,

We thank the reviewers for a careful reading of the manuscript and their suggestions. The manuscript has been improved as all raised concerns have been addressed. The edited parts of the manuscript are highlighted in bold face.

Thank you for your consideration,

The authors

Reviewer 1 (Patricio Orio)

Basic reporting

Although the findings are interesting and well sustained, my major concern is with their novelty. Diversity in neural networks has indeed not heavily studied, and its role to network function remains to be clarified, but there have been some recent works by other authors, more than what is mentioned in the paper (In other words the issue has been studied more than what the authors want the readers to think). In particular, I find it difficult to see what is novel in the present work compared to that of Mejias & Longtin (2014), already cited in the paper. Although the models are different (mainly continuous versus discrete) the conclusions are at first glimpse very similar: heterogeneity flattens the f-I curve, meaning a larger dynamic range. If there is some novelty in the details of the results or the analyses, then this should clearly stated and thoroughly discussed. Also, it should be clearly stated why it is worth revisiting an apparently settled issue with a discrete model approach.

Although the general topic has already been studied, our approach is very different, and our results can address questions that remained untouched such as: what makes a certain level of diversity optimal? In the simple bimodal distribution we identify tricriticality to be a crucial factor to generate optimal network response. Additionally, we study different (more complex and likely more realistic) scenarios. We have edited the introduction on pages 2 and 3 and the discussion on pages 13 and 14 to highlight the novelty of our contribution.

Mejias & Longtin (2014) do indeed find that diversity affects the f-I curves. However, we go beyond their work by characterizing how this influences the coding performance of networks to external driving that varies over orders of magnitude. We attempt to identify what causes the observed optimal level of diversity; we focus on subpopulations within excitatory/inhibitory neurons; and we vary the distribution of thresholds. These are all important elements and the last one, for example, has triggered a commentary paper explicitly drawing attention to this problem (reference [56] F. Baroni and A. Mazzoni, Heterogeneity of heterogeneities in neuronal networks," Frontiers in Computational Neuroscience 8 (2014)). We address these elements, and make it clearer on pages 2, 3 and 13.

Some other works to be cited and discussed are: Hunsberger et al., Neural Comput 2014; Savard et al., Neuroscience 2011; Metzen & Chacron, J Neurosci 2015, Mejias & Longtin, Phys Rev Lett 2012.

We now cite and discuss these references on page 2.

The Mean-Field results are poorly described in the Results section, and there is no actual conclusion drawn from them. A better description of the mean-field approach as well as a comparison of its results with the simulation results are needed. Otherwise I suggest these results to be removed from the Figures and Methods section.

The mean-field approach is described more thoroughly on pages 4 and 5; the results of the mean-field approximation our now discussed as the results are presented on pages 5, 6 and 7. Please note that black solid lines in figures 3, 4 and 5 correspond to mean-field results, whereas symbols represent simulations.

Raw data has not been made available. For this type of work, I expect model codes to be deposited in an appropriate repository.

C code to simulate the model is now available on our lab's website at www.sng.org.au/Downloads.

Experimental design

No comments

Validity of the findings

* Line 242 says: “(iv) The collective response from the entire network can outperform all subpopulations but only when the specialized subpopulation is far from its optimal tuning”. I don't see this clearly. From most of the results, it appeared to me that the network outperforms non-integrators depending on the distribution of thresholds rather than the unit's tuning. Please clarify.

This has been corrected (see first paragraph of the Discussion, page 11).

* Raw data has not been made available. For this type of work, I expect model codes to be deposited in an appropriate repository.

C code to simulate the model is now available on our lab's website at www.sng.org.au/Downloads.

Comments for the author

Some minor comments:

* The critical value \lambda_c, mentioned for the first time in line 122, should be marked in Figure 2a with an arrow or similar symbol.

This has been added to Figure 2a.

* The susceptibility (line 126) must be explained in plain words, similar to how the dynamic range is explained.

We now explain susceptibility (now on line 153) in addition to providing the mathematical definition: The susceptibility captures the variability of the instantaneous ensemble firing rate around its mean value (over time) for each subpopulation.

* The concept of percolation should also be explained and its importance discussed.

Following the suggestion, we have now extended our discussion on the topic of percolation on pages 13 and 14.

* In Figure 8a, only the specific case in which the network outperforms the subpopulations is shown. An additional figure, showing the most relevant case of the non-integrators outperforming the network, could also be shown.

We added two additional panels in Figure 8 to illustrate with more detail the different dynamic regimes. Figure 8 now compares a cases in which non-integrators outperform the network (Figure 8 d), a case in which the network outperforms the subpopulations (Figure 8 f), and a third intermediary case in which the parameters describing the gamma distribution is located between the two previous cases (Figure 8 e).

* The sub-subsection of Networks with Inhibitory nodes (lines 213) should be a section with a higher hierarchy, i.e., similar to the other sections. In other words, the two sub-subsections under 'More realistic scenarios' are too different to be considered as being part of one subsection.

We have divided this section.

* In the networks with inhibitory nodes it is said that 20% of units are inhibitory. In the heterogeneous network, how is this 20% distributed among the different threshold populations? 20% of each?

We introduced bimodal diversity in the thresholds as follows. First, we fix the proportion of inhibitory units at 20%. For each total density of integrators, we distribute these according to three simple cases: (i) all inhibitory units are integrators (thus requiring a total integrator density d>=20%, with the excitatory units comprising the d-0.2 integrators and the remainder nonintegrators); (ii) all inhibitory units are nonintegrators (thus requiring a total integrator density d<= 80%); and (iii) diversity in the threshold of the inhibitory units (fixed at 50% integrators and 50% nonintegrators, thus requiring a total integrator density 10% >= d <= 90%). This covers the two extreme cases (i) and (ii) and an intermediate case (iii). Their similarity emphasizes the point that inhibition does not play a major role in shaping the curves of the dynamic range. This section has been rewritten to clarify this point.

* In the first lines of the section describing networks with inhibition (lines 214-222), it s not completely clear which statements are findings of the present work and which are previous findings. The fact that the inhibitory mechanism of the model is explained later (starts in line 222), made me think that all the previous lines are all previous evidence. Is this correct? Please add some more references or change the writing style to avoid confusion.

The reviewer is correct; we have restructured the section and rewritten the paragraphs to clarify this point (page 10).

Reviewer 2 (Anonymous)

Basic reporting

The article is very well written, easy to follow and understand, with a thorough literature review. The figures are also very clear, complement the text and are also easy to understand. As it is, the only defect I find in this section is the unavailability of the networks used in the study.

Experimental design

With respect to the experimental design, I found several flaws that require further attention.
1.- The authors do not specify if they used the same network architecture (connectivity between neurons) and only altered the types of neurons or if they generated a random network for each network with different composition.

We have clarified on page 4 that we use an ensemble of random networks, one for each trial.

2.- It is also not specified how many times did the authors generated a random network or if they did it only once. This is an important aspect since different connectivity between nodes can produce different network behavior. In other words, are the results shown in the article the product of averaged performances from several replica of the networks or obtained from the same single random network?

Each trial uses a different random network. We have also added text to explain that " Although each network exhibits its own distinct dynamics, the ensemble average responses are very similar across trials."

3.- In lines 71-72 in page 3, the authors claim their results generalize to other sizes, connectivities and topologies. This strong affirmation is not proved or demonstrated anywhere, and making it should require prove, at least from my point of view. To prove it, several replica of different random network models generated with different approaches should be tested together with different network sizes and connectivities.

The statement is correct but this is not a main point of the manuscript. Our intention was not to prove it but rather to clarify that network specificity does not play a major role in our results. The addition of this extra material would make the paper too long and dissolve the main message. This part has been rewritten and the claim deleted.

Validity of the findings

Even the results are sound, without addressing the flaws in the experimental design, they lack validity. This is highly relevant specially with respect to the statistical significance of the reported findings, since without several replica of the random network generation it is not possible to calculate.

As clarified above in response to points 1 and 2, we used an ensemble of random networks. The results correspond to an average over several trials with different network topologies. This is explained on page 4 of the manuscript.