Dear editor,

Please find below our detailed response to the reviewers’ comments for our manuscript entitled “Cell type specific firing patterns in a V1 cortical column model depend on feedforward and feedback input”, currently under consideration at PLoS Computational Biology. We have also provided, with this submission, a copy of the manuscript with highlighted changes and a fresh copy with changes included but not highlighted.

We have carefully considered the suggestions of both reviewers and have profoundly revised our manuscript, with a particular emphasis on improving several modeling choices to make our model more realistic. As a consequence of that, we have (in brief):

* improved the core computational model (which is now introduced for the first time in this manuscript, instead of referring to our manuscript ref. 21), in particular the way we introduce connectivity data and details on implementing the external input to the model,
* included a new section about the statistics of the model in the spontaneous state,
* Included new analysis such as (i) a finite-size study to confirm the validity of our model for larger system sizes, (ii) perturbation matrices with more information about the effects of perturbations and the use of the Frobenius norm for cross-comparisons, (iii) a study of the effects of different perturbation strengths, and (iv) power spectra and irregularity metrics for all conditions.
* updated all main figures with the improved model and included a new one (Fig. 2),
* updated all supplementary figures and added 11 new ones, for a total of 14,
* carefully addressed all reviewers’ comments in the text.

A detailed response to the comments of each reviewer can be found below.

We are happy to report that the main message of our study, along with the main results presented in the previous version, remain valid after including the improvements suggested by the reviewers, and the new analyses provide a. Therefore, we hope this new version may now be considered suitable for publication in PLoS Computational Biology.

Yours sincerely,

Giulia Moreni, Cyriel Pennartz and Jorge Mejias

**Reviewers’ comments (in black) and response by the authors (in blue):**

**Reviewer #1**: In this manuscript, the authors report an application of a model of a “cortical column” in mouse V1 that they previously developed. The 5,000-neuron network, using point neuron models of diverse types and explicit representations of cortical layers, was built following previous studies and data from the literature. Here it is simulated with feedforward (FF) and feedback (FB) inputs. The authors apply perturbations systematically to every cell type in the model one by one and investigate the effects of these perturbations on different cell types and under different conditions (spontaneous activity vs. FF-evoked vs. FB-evoked). They discuss a number of findings focusing on the distinct ways of how the network reacts to the perturbations under these different conditions.  
  
 While overall the approach is interesting and the results could be useful, unfortunately there are many issues with how the model was constructed and simulations executed. Too many choices appear to be not biologically realistic, which decreases the value of this work. The study would greatly benefit from addressing these issues and revising simulations accordingly.

Our response: We thank the reviewer for his/her feedback and especially for carefully revising the validity of the model’s assumptions. Although it is not our intention here to provide a highly realistic cortical column model, but rather to develop an adequate (yet simplified) description to explore the state-dependent effects of perturbations in columnar dynamics, we fully agree on the importance of proper biological approximations. We therefore have taken these concerns very seriously and, after revisiting all our model choices, we have improved the model in several ways and redone all simulations (besides adding other extra analyses).   
  
1) The Methods section does not provide the connection probabilities, instead pointing to Moreni et al., which, in turn, points to Billeh et al. Following that trail, one finds that the connection probabilities numbers reported by Billeh et al. are in fact levels of probability for a distance-dependent Gaussian function p(r) at r = 75 um. That is very different from how Moreni et al. use those numbers – simply as overall connection probabilities, which is incorrect. The authors should revise their approach to these data. There is some not too complicated math that can be used to translate Billeh et al.’s p(r=75 um) numbers to approximate values of uniform connection probabilities. The result depends on the radius within which the network is defined and the width of the Gaussian, sigma. A back-of-the-envelope calculation suggests that the correct uniform probabilities should be somewhere in the range of a factor 1.2-1.6 higher than what Moreni et al. used, depending on the sigma, which varies depending on the connection type. This is a substantial difference. The authors should revise their model with correct connectivity numbers and check what results they get.

Our response: We agree with the reviewer’s point, and we have now corrected the way in which we introduce the connections probabilities following his/her suggestion.

With a realistic cell density, a network of N=5,000 neurons would correspond to a cortical column with radius R ~= 125 um. Our calculations, using this radius, shows that the connection probabilities are indeed slightly different from the ones used in the first version of the model. In the revised version, we have a probability for each cell type depending on the distribution’s width (i.e. sigma), which varies depending on the connection type. These new probabilities are the ones we now use in all results along the revised version of the manuscript.

The calculations, as now appear in the new description in the Methods section, explains the procedure:

Since, for realistic cell densities, the connection probabilities in Billeh et al. correspond to a column with radius of r=75 um and our own network of 5,000 neurons would correspond to a column of radius ~125um, we translated those quantities to better approximate values of uniform connection probabilities.

We performed the translation for each connection type using the correct value of the Gaussian width of each connection, denoted as sigma. These values of sigma are reported in supplementary material of Billeh et al. and also in the following table.

The corrected uniform probabilities used in the model are given by

where, is the probability used previously, R=125um, = 75 um, and sigmas are reported in the following table along with the result of the calculation.

|  |  |  |
| --- | --- | --- |
|  |  | *New probability* |
|  | 114 *um* | \* 0.716 |
|  | 92 *um* | \* 0.6244 |
|  | 103 *um* | \* 0.6742 |
|  | 95 *um* | \* 0.6388 |
|  | 85 *um* | \* 0.589 |
|  | 120 *um* | \* 0.7371 |

2) The model contains 5,000 neurons total. This is a small number, and it is possible that conclusions from simulations will change substantially if a larger system is used. The reason is that connectivity is far from being saturated with such a small number of neurons. That is particularly bad for the connectivity model used by Moreni et al., since it is not distance-dependent, so, any increase in the number of cells in the network simply increases the number of connections per cell proportionally (if N increases by a factor of 2, the number of connections a cell receives from other neurons increases by a factor of 2). But let us assume a less bad case, where connectivity falls off with distance (as it does in the cortex). With realistic cell density, N=5,000 neurons correspond to a cortical column with radius R ~= 125 um. Given a Gaussian connectivity function, one can then compute how many cells of a given type, on average, will receive input from (and provide outputs to) a cell at the center. What if the system had 4 times more neurons, N=20,000, so that R is twice larger, 250 um? Assuming Gaussian sigma = 110-120 um (a typical value for the mouse cortex), one finds the number of connections for the cell at the center grows by a factor of a bit over 2. If we keep increasing N, and therefore the radius, that factor keeps growing and saturates at approximately 2.5 or so. Note, this is for a cell at the center. Cells at the edge of the system have many fewer connections, and the effect on them will be far more dramatic. What this all means is that the current model very substantially underestimates the number of connections in the system, that is, the connectivity is much sparser than it should be. That probably makes it much easier to have a stable network, but very strongly distorts the dynamics, which for a more realistic network would be dominated much more substantially by recurrent connections. This is a serious issue, and the authors should address it either by running all their simulations with a larger model (perhaps N=50,000 or so) or at least test key findings with such a larger model.

Our response: We agree that, in general terms, results from simulations of small networks may deviate from those of larger networks closer to the real size of cortical networks. However, we would argue here that the focus of this work is on changes in mean firing rates, which typically do not present severe scaling problems (especially when appropriately scaling synaptic weights like the reviewer also acknowledges in the following comment). For example, scaling studies by van Albada et al. (PLoS Comput. Biology 2015) show that, in cortical column models similar to our own, scaling has a considerable effect in cross-correlations, but the effect on mean firing rates is comparatively smaller and can be addressed with proper synaptic scaling. This is an important consideration since, due to the exploratory nature of our work (which involves computing perturbation matrices across many conditions), having to simulate full-sized cortical columns would make the study prohibitive in terms of computational time.

To make sure our network model is able to estimate firing rates of larger networks accurately, we performed a finite-size analysis in which we compute the spontaneous firing rates of all layers and cell types across a wide range of network size (5k to 80k). The result (see new Fig. S2 and new paragraph in page 6) shows that our network of 5k neurons was able to provide good estimations of the activity levels obtained for larger networks (up to 80k), with variations of ~1 spike/s or less. Therefore, it is reasonable to assume that our results, which are based on firing rate values, will hold for large networks.

3) I realize that the authors used scaling of synaptic weights to address this issue, per Eq. (11) in Methods: w\_j = G \* s / (N\_send \* p). But that seems problematic as well. As the authors write, “s is taken from the experimental synaptic connectivity matrix S” from Billeh et al. However, that matrix in Billeh et al. is very clearly a matrix of unitary PSPs in mV. That is, “s” is how much of a PSP is created, on average, in the target cell by a single spike of a source cell, in millivolts. Take, for example, the E-to-E connections in L2/3. According to Moreni et al. Eq. (11), w\_j = 5 \* 0.36 / (1236 \* 0.16) = 0.009. Is that then the weight for each connection of that type, 0.009 mV? That is much smaller than Billeh et al.’s w\_j = 0.36 mV. And that small value is used in the network that already has much sparser connections than in the actual biological V1 network, as discussed above. Again, that would lead to a much weaker effect of recurrent connectivity on the network dynamics than warranted by data. I believe the authors should revise their models and use the appropriate weights, so that PSPs from each individual connection are on average consisted with the S matrix of Billeh et al., if that is the source of data they like to use. It will be interesting to see if the results of simulations change or not.

Our response: We think the reviewer might be mistaken here, as the value of w\_j does not correspond to the PSP in our model. Indeed, w\_j is a dimensionless quantity in our model, and the overall value of the PSPs is given by the outcome of Eqs. 3-6 when introduced in Eq. 1. Therefore, the concrete value of the PSPs depends on the presynaptic spike history (especially since we have NMDA contributions which introduce long time scales) and the distance of V(t) from the reversal potentials. In practice, PSPs in our model vary considerably over a range of about 2-3 orders of magnitude which includes the values observed by Billeh et a. It is however not the goal of this model to provide accurate predictions of PSPs for each cell type (especially since this is known to vary significantly depending on brain state, ongoing or past learning processes, and other factors) but rather to generate dynamics consistent with observed firing patterns at the population level. We have clarified this in a new sentence below Eq. 6.  
  
4) Another related issue is that all connections of a given type in Moreni et al.’s network have the same weight. That is, of course, unrealistic – study after study reported that PSPs or PSCs of synaptic connections in the cortex are distributed over a broad range, typically 1-2 orders of magnitude at least.

Our response: We agree with the reviewer that the absence of PSPs is an important limitation of our model. However, this is a natural limitation also present in similar models in the literature and, to our knowledge, there is no dataset of enough quality to overcome this issue for the full cortical column with four cell types and AMPA/NMDA/GABA synapses. As we said in our previous response, it is also not our aim to address the issue of predicting realistic PSPs values specifically, given the experimental variability in such numbers. A more organic way to introduce PSPs would be to consider the introduction of synaptic plasticity in the model. We are currently working in this direction, but the implications of such extra modelling work fall well beyond the scope of this paper. We have added a small mention in a new sentence below Eq. 6.

5) The manuscript repeatedly cites Billeh et al., Neuron, 2020 with regard to choosing model parameters (cell types, cell properties, connectivity, etc.) and referring to that as “anatomical data” and other such terms. However, Billeh et al. reported a model – it is not a primary source of experimental data. That in itself is fine but calling it “anatomical data” is misleading. For example, Billeh et al. used a variety of different papers to inform their connectivity choices; moreover, their connection weights were heavily optimized to achieve a model with desired dynamical properties. Therefore, using parameters from that model is not the same as taking them directly from data. That should be acknowledged, perhaps simply by stating that the model of Moreni is based to some degree on the model of Billeh et al.

Our response: We agree with the reviewer and we have revised the full text to acknowledge Billeh et al.’s contributions to our work in terms of connectivity, modeling and parameters instead of “anatomical data”. The neuron parameters, such as membrane potential threshold, capacitance etc. were indeed taken from Allen database, as explained in Method section “Parameters of the model”.  
  
6) Equations (3-6) in Methods: I cannot find anywhere a description and/or values of the conductances g\_AMPA, g\_GABA, and g\_NMDA.

Our response: We have properly described the parameters g\_AMPA/GABA/NMDA below Eq. 6. They constitute simple conversion factors which are given values of 1 nS, as it is now stated in the text. Note that this doesn’t mean that all three synaptic types have the same strength, as other parameter also contribute to this. Indeed, NMDA contributions are weaker (but more prolonged) than those from AMPA, and GABA contributions are strong enough to maintain an inhibition-stabilized regime (those strength depends, among other things, on the global coupling G). The relevance of these ratios fall beyond the scope of this work, but will be addressed in future studies (for example, focused on E-I balance).  
  
7) The I\_ext(t) is not described. Is this the current injection to simulate perturbation?

Our response: I\_ext(t) is an external input to the neurons, for example is the FF input or FB input, and provides the ‘state’ of the network. This is mentioned in the paragraph right above Eq. 2, has been highlighted more now. Perturbations are instead simulated by injecting a perturbative current to the target cell type and this is indeed accounted for in the I\_ext(t) variable.   
  
8) It appears that FF and FB inputs are modeled as current injections (so, I\_ext(t) above, like for the perturbations)? That should be explained clearly in the Methods. But also, this is not a great choice - FF and FB projections very clearly provide synaptic inputs to V1 cells, and it would be much more appropriate to model them as such.

Our response: We have provided a more detailed explanation now in the Methods section as the reviewer asked, under the subsection “Perturbation matrix”. Regarding the approach to modeling these contributions, we rely on the diffusion approximation which is used in seminal work by Brunel and others, so we decided to use this approximation in our model. We have added a paragraph explaining this in the Methods section, paragraph “Perturbation matrix”.  
  
9) A substantial limitation of the study is that every perturbation is simulated as a 30 pA current injected to every cell of the perturbed type. The effects are going to differ substantially depending on the magnitude of such current injection, therefore necessitating sampling different magnitudes, from 0 pA to 30 pA and well beyond. Both positive and negative injections should be considered, because perturbations that shut down a neural population are perhaps even more common experimentally than those that stimulate. Moreover, 30 pA means very different things for different cell types, depending on their input resistances and rheobase. Sampling multiple injection magnitude for each cell type could help with that, but ideally the authors should calibrate the perturbations separately for each cell type.

Our response: The reviewer is correct, the perturbation effects change depending on the magnitude of the perturbation. To address this point, which is similar to another one raised by reviewer #2, we repeated the analysis (i.e. the 16x16 perturbation matrices across the spontaneous, FF, FB and FF+FB conditions) with a broad range of perturbative inputs, both positive and negative. More precisely, the perturbation analysis in each condition is now run with the following range of perturbation values : [-40, -30, -20, -10, 10, 20, 30, 40] pA. In Fig. S12 we show, for each condition, the number of changes in firing rates bigger than a given threshold introduced by the perturbation. As one would expect, the number of changes increases with the strength of the perturbation (in absolute value) and decreases as we increase the threshold of change with respect to baseline (%). Likewise, the positive and negative branches present slight asymmetries (as the effect of inhibitory perturbations is bounded by the fact that firing rates must remain positive). Besides this scaling, our results remain the same. We have included a brief description in page 22, right before the Discussion section.  
  
10) For FF inputs, the only description I see is “input to all pyramidal neurons in layer 4”. That is an oversimplification. It is questionable whether all L4 neurons in V1 receive FF thalamocortical inputs. More importantly, it is well known that FF thalamocortical inputs project diffusely into mouse V1, and certainly not exclusively in L4 (though projections to L4 are the strongest). There is evidence that FF inputs target inhibitory neurons, and not only pyramidal neurons. In particular, their targeting of PV interneurons – feedforward inhibition – is considered an important mechanism in gain control and information processing. The fact that this mechanism is absent from Moreni et al.’s model makes the model quite unrealistic – feedforward inhibition certainly has a strong effect on the network dynamics. Furthermore, even for the strongest visual stimuli, by far not all thalamocortical inputs are activated. Typically, only a small fraction of them will be firing in response to any particular visual stimulus – those thalamic neurons that are tuned to that stimulus. Simulating FF input as a strong activation of all FF projections to V1 is inappropriate – it never happens in real animals.

Our response: While it is true that FF inputs are somehow diffuse across targeting all layers, we chose to target only layer 4 to simplify the description of the model. However, we agree that other assumptions, including targeting all pyramidal cells in layer 4 (and only pyramidal cells) might have important consequences in the results and make our simulations less realistic. Therefore, we followed the reviewer’s suggestion and we redid all our simulations in the study with a FF input which targets a subset (25%) of layer 4 pyramidal neurons, and also a small subset (5%) of PV cells (for higher percentages of PV activation, such as 20%, overall firing rate changes are small and not compatible with experimental observations). Similar E-PV percentages are used when we consider FB input into different layers of the column. We now explain this in page 37.

11) “These stimulus-triggered firing rate responses are in line with experimental observations in vivo [21].”  
  
 First of all, the firing rates are not shown. The statistics of the rates should be shown for all major conditions, such as spontaneous, FF-, and FB-evoked, and compared to experimental vales. By the way, we do not really know what FF- and FB-evoked rates are on their own – the experimental data are mostly available for the situation where both FF and FB inputs are present. Second, ref. [21] is the original model by the same authors, and it does show some firing rates, but as far as I can tell, it compares the rates to Billeh et al., which is a simulation study, not an experiment. And then, importantly, the rates observed in response to FF stimulation are way too high: e.g., 15 Hz on average for excitatory cells in L4. In real mouse V1, for any visual stimulus, a small fraction of V1 neurons that are tuned to that stimulus will be activated, with median firing rates of ~5 Hz or so for excitatory cells, for their preferred stimulus. Most neurons that are not tuned to that stimulus remain close to baseline. Thus, even 5 Hz on average for all E neurons in L4 would be way too much. But 15 Hz is totally unheard of. All that is to say that the authors apparently simulate a very, very strong activation of V1, which is not a normal regime in which this circuit operates in vivo. It would be great to revise the model and simulations accordingly.

Our response: We have now removed this sentence, as we were indeed not showing these results explicitly in this work and we don’t have experimental data to compare directly with our simulations for all conditions (as the reviewer correctly points out, this is difficult data to obtain).

Additionally, with the new modeling choices for external input (targeting E and PV cells, and only a subset of neurons within each layer), our firing rates are now comparable or at least closer to vivo conditions (i.e. from spontaneous levels up to a maximum of 10 spikes/s). The only exceptions would be pyramidal layers 4 and 5 during FF+FB activity, which rises up to 15~20 spikes/s. However, we do not agree with the reviewer about levels of 15 Hz being “totally unheard of”. To mention just one example, firing rates of up to 30 spikes/s have been found in mouse V1 (Niell and Stryker, J. Neurosci. 2008, Fig. 2A). While our model doesn’t include orientation selectivity properties like the ones reported in their experimental study, this constitutes proof that, physiologically speaking, these firing rates exist in mouse V1. Future work in our columnar model, for example introducing orientation selectivity profiles, could indeed reveal that our levels of 15~20 spikes/s can only occur for highly preferred input.

For completeness of our study, we have performed analysis of the statistics of the spontaneous condition of our model (as also requested by reviewer #2) and compared the firing rate levels with available experimental values (new Fig. 2). Statistics for the other conditions (FF, FB, FF+FB), for which experimental data is not directly available, have been obtained as well with our model to facilitate future comparisons with data (Fig. S4, S13 and S14) and fall within reasonable levels of irregular spiking patterns.

12) “feedback signals corresponding to top-down modulations from higher cortical areas can be simulated by injecting external current into layer 5 pyramidal neurons [17, 19, 29, 30].”  
  
 As the authors point out later themselves, that is not the best choice for feedback to mouse V1 (and refs. 17, 19, and 29 are for primate cortex). It is known that feedback targets at least L2/3, L5, and L6 in mouse V1. It is great to see that the authors addressed multiple scenarios of where FB might be coming in. However, they did not address the one scenario that makes the most sense to me, which is FB coming into L2/3, L5, and L6 at the same time. I would make that a default scenario (rather than FB to L5 only) and consider other scenarios later in the Results.  
  
And like for the FF inputs, activating all FB inputs at once is unrealistic. For any visual stimulus, a small fraction of them will be activated.

Our response: Although, for simplicity reasons, we prefer to maintain the input to layer 5 as a default case for FB condition, we fully agree with the reviewer about the importance of the FB input targeting L2/3, L5 and L6 and we have now performed new simulations in this new scenario. We show these results now in Fig. 5. Note that we have also changed the number of targeted cells also in the FB conditions, so now only 25% of excitatory and 5% of PV cells are targeted.

13) Fig. 2 and later figures, as well as text: the notation “> -20%” is very confusing. The authors meant that these are situations where firing rate is reduced by 20% or more, right? If so, just denote it as “< -20%”. I also do not understand why it is necessary to plot the matrices using such thresholding for the colors. They can simply use saturation of blue and red to denote the percentage change. With thresholding using an arbitrary number of +/-20%, the figures are somewhat misleading about the actual simulation results.

Our response: We agree with the reviewer and apologize for the confusion, we meant a situation where firing rates are reduced by more than 20%. We changed the notation now to <-20% as the reviewer suggested.

Regarding the color code, and in response to a similar comment by reviewer #2, we now included the complete colour maps in supplementary material for each condition denoting the exact percentage change without a cut at 20% (see Figs. S5, S6, S9 and S10). In the main figure we kept the 20% matrix because it can give a better intuition of the bigger changes, and provide the more complete data in the supplementary for full disclosure of all changes.  
  
14) “However, a multi-compartment approach adds complexity, potentially overcomplicating the model, making it more challenging to manage. This is why multicompartments were not included in the current model, as they likely wouldn't significantly affect the results concerning overall perturbation dynamics.”  
  
 That last sentence is strange. There is no data shown to suggest that results would not be significantly affected. I would recommend to remove it, as it does not add anything.  
  
 The whole paragraph before that describes the phenomena that might be associated with distributed processing in more realistic multicompartmental neuronal models and rightfully mentions the absence of such models in the present study as a limitation. We don’t know if it makes a difference until we try.

Our response: We fully agree with the reviewer, and we have now removed this sentence from the text.  
  
15) “An explicit study of natural stimuli would require mapping visual inputs directly to layer 4 neurons” – as mentioned above, the current study could be substantially improved even without a comprehensive representation of the pixel-wise visual input. One should simply reduce the number of FF inputs activated at once and limit each activation to a small subset of cells. That is much more realistic than a massive barrage of inputs into all L4 pyramidal cells.

Our response: We agree with this suggestion, and as stated in our response to comment #10, we have now changed all the simulations by giving input to only a subset of neurons. In the case of FF input, for example, only 25% of excitatory cells and 5% of PV cells in layer 4 are targeted.   
  
16) What software was used to build the model and run simulations?

Our response: We have used custom Python code and Brian2. This is now mentioned in the Data and materials availability section.  
  
17) It is great that the authors promise to make all code available on GitHub upon publication (although an even better approach is to make it available before publication, at least for reviewers). However, are model files and simulation results available or will be made available? If not, they should be.

Our response: Our code has been now uploaded as a zip file for reviewers. As the reviewer suggests, the code, along with model files and simulation results, will be made available upon acceptance.

**Reviewer #2**: In this paper, the authors present a very extensive numerical experiment, checking the response to simulated cortical columns under perturbations. The study is well designed and the methods seem to be robust. Although I would not call it very impactful, I believe that the content presented in this paper can be of interest to both computational neuroscientists and experimentalists. It's a simple, classical model, but the authors made an effort to make very extensive comparations with realistic parameters. The paper is straightforward to read.

Our response: We thank the reviewer for his positive evaluation and careful analysis of our work. We have addressed all comments below.

Under this point of view, I have only few major issues that I would like to address:  
  
 1) The authors quantify the effect of perturbations by identifying differences of +/- 20% in baseline firing rates, constructing "response matrices". Among those, one of the most important response matrices is the one obtained in Fig 2C (and similar ones through the paper), where authors look for differences between the perturbations in the spontaneous case and the condition under study. My main concern abuot this is that authors talk about the "significance" of the differences, which seems to suggest that the white squares in matrix in Fig 2C are not \*statistically significant\*. While it is true that the changes in baseline firing rates can be low, they can still be statistically significant, as measured by appropiate tests. Therefore, I would like the authors to conduct statistical tests and plot a matrix, as figure 2C, but in which empty elements actually show which changes are statistically not sigficant.  
  
 If the results are still similar enough to the ones obtained by the authors, I would propose to use statistical analysis instead of the current plot through the paper, as it has no possible confusion and it is stronger. However, I do believe that many "non significant" changes will suddenly become statistically significant. I do not think that this invalidates previous results from the authors, but the language of the paper will have to be updated (to avoid talking about significance, and focus on which perturbations lead to larger changes in rate) and the discussion will have to be updated in the light of the new results.

Our response: We apologize for the confusion. Indeed we didn’t mean “statistically significant”, as classical statistical analysis of significance will not very informative for our computational model –upon enough simulation trials, all changes will be statistically significant. We meant to show instead whether the change of firing rate introduced by the perturbation was above or below 20% of the original value.

We have carefully revised the text and avoided any mention to the significance of the results, to avoid confusion by the readers. Moreover, we have now provided a more detailed matrix, also requested from reviewer #1, to provide more information about the effects of perturbations in a clearer manner, for all perturbation matrices and conditions computed in our study (see also our response to comment 5 below). With our new analysis (see Figs. S5, S6, S9 and S10) we now show the exact percentage changes in firing rates without cutting them with a 20% threshold. We hope this additional information provides more clarity to our results.

2) The authors keep the perturbation strength constant for all the experiments. However, being the system is highly non-linear and showing a non-trivial connectivity it is possible that different perturbation stregngths lead to very different response matrices. I would ask the authors to ensure that the results are consistent for a reasonable range of the perturbation parameter.  
  
 In the same line, the authors might would like to compute and show linear response of the system, i.e., to compute the derivative of the stationary firing rates as a function of the perturbation, dr/dh. Populations with larger linear response will be the ones having larger responses to small perturbations, which is what the authors are showing, while avoiding the pitfall of selecting an "adequate" perturbation value.

Our response: We agree with the reviewer and, following also a suggestion by reviewer #1, we have now addressed the effect of different strengths of perturbations in our results. To be more systematic, we have opted for the first suggestion of the reviewer (also highlighted by reviewer #1) and explored our model’s response for different perturbation values. The perturbation analysis in each condition (Spontaneous, Feedforward, Feedback and Feedforward + Feedback) is run with the following range of perturbation values : [-40, -30, -20, -10, 10, 20, 30, 40] pA. We observe that the main results are still present, i.e. the responses are dependent on the state (Fig. S12). As one would expect, the number of changes increases with the strength of the perturbation (in absolute value) and decreases as we increase the threshold of change with respect to baseline (%). Likewise, the positive and negative branches present slight asymmetries (as the effect of inhibitory perturbations is bounded by the fact that firing rates must remain positive). Besides this scaling, our results remain the same. We have included a brief description in page 22, right before the Discussion section.  
  
 3) It is implicitly assumed that for all experiments and conditions, firing rates are always constant, i.e., there are no other dynamical regimes such as oscillations. The authors should control that this is the case throughout the paper (since inhibition alone is able to generate oscillations) because the measurements done could have little sense otherwise. Moreover, if dynamical regime changes under perturbations it would give useful insights.

Our response: This is indeed an important point. To control for the existence of oscillations, we have conducted a power spectrum analysis for all different conditions considered. In Fig. S3 we show the power spectrum for excitatory neurons in the different layers. We didn’t find oscillations in any of the different regimes. For the case of feedforward plus feedback input, the neurons receiving the external input show a slight level of synchronisation (as reflected in a small peak in the power spectrum), but such modulation has fast frequency and very weak and inconsistent amplitude, and therefore it is not able to introduce alterations (such as phase-dependent input responses) in our results.

4) I am missing some controls on the dynamics of the initial states, especially in the spontaneous one. The authors talk about balance in the paper, but I believe that authors do not try to impose any kind of balance in the system. Is this obtained automatically from the experimental parameters given to the model? Is balance altered in the feedforward of feedback conditions? Is the dynamical state of the spontaneous system irregular, i.e. with CV~1? Again, how does the picture change under alterations? I would appreciate if the authors could add a supplementary figure to ensure correctness of the simulations, showing that indeed the spontaneous dynamics is an irregular state.

Our response: The reviewer is correct, the previous version of the manuscript was missing some important information about the properties of the spontaneous state, and in particular about its balance. Regarding the information on the spontaneous state, we have now included a new figure (Fig. 2, accompanied by a new section in the main text) to provide more information about the spontaneous state. As the new figure show, the spontaneous state shows overall spiking patterns which are irregular and of low synchrony, and with activity levels that are compatible with experimental recording across all layers and cell types.

Regarding the E-I balance, it indeed obtained automatically from the experimental parameters of the model (and provided that the global coupling G is strong enough). To demonstrate that, we have added a new supplementary figure (Fig. S1) which shows that suppressing inhibitory input within the network leads to an explosion of excitatory activity, suggesting that the original, unperturbed network is in a balanced state with inhibition providing overall stability to the network. In other states (FF, FB, and FF+FB), balance is initially altered by the external drive but inhibitory neurons are able to compensate for it, as the consistent high levels of irregular activity (CV) for those conditions reveal (Fig. S4, see also S13 and S14). The only exception are pyramidal neurons in layer 2/3 (which display very low CV), but the fact that their firing rate level is also very low (0.27 spikes/s) makes the numerical estimation of the CV more unreliable in this case.

Then, some minor issues follow:  
  
 5) I would plot figs 2abc with a continuous color scale. I understand the authors want to highlight only significant changes, but I think at some point the whole matrix should be colored in an easily legible colormap, as it might be of help for experimentalists willing to check their experiments against your predictions. These could be added as sup figs.

Our response: Thank you for the suggestion, this was also indicated by reviewer #1 and motivated us to add fully color-coded response matrices as supplementary figures for all and each condition in the manuscript, showing the exact percentage change in firing rates (see also our response to comment #1 above).  
  
 6) Equations would benefit from indicating explicitly that quantities are referred to a single neuron, so for example currents would have an \_i subscript, and multiplication by the adjacency matrix would be by sum\_j w\_{ij} in equations 4,5.

Our response: We understand the reviewer’s concern about clarity of the equations. We would prefer, however, to maintain our current style, as we aim to follow the nomenclature from classic works (Wang 2002) and would like to maintain coherence with those works.

7) The authors say that response matrices in Fig. 4 look very similar. They could quantify the differences by an appropiate matrix norm, such as Frobenius norm. As an idea, authors might compute pairwise distances between all the matrices and see which conditions are closer to each other.

Our response: We have followed this advice and we have now introduced the Frobenius norm and computed the distance between all response matrices considered (see Methods section Application to Matrix Similarity, and Fig. S9)

8) In equation 11, I would write a dot or whitespace, as the star (\*) could also be convolution (and the "hat" symbol is often use for Fourier transform, thus inducing some readers to think that authors are doing some Fourier-space maths).

Our response: Agree, we changed this accordingly.   
  
 9) It is great to have the exact connectivity matrix in Fig 1, but it is difficult to read intuitively. Adding the same information in a network-like sketch might be useful for the readers.

Our response: We have now added a wiring diagram in Fig. 1 to make the connectivity structure more explicit and intuitive.  
  
 To sum up, I think the paper is sound and could be a useful reference. Hence, I am in favor of endorsing publication in Plos Comp Bio, as long as the major issues are satisfactorily answered and no other issues emerge in the revised version.

Our response: Thank you again for the positive evaluation. As our significant revision and updating of the model has provided the same results, we hope that this serves to reassure the reviewer about the robustness of our results.