Point-by-point reply to the reviewer's comments

Manuscript No. eN-TNWR-0016-23R1

"Stable neural population dynamics in the regression subspace for continuous and categorical task parameters in monkeys"

We greatly appreciate the many efforts from the reviewers and editor. Below, we provide point-by-point responses to the remaining reviewers' comments.

Reviewer comments:

Thank you authors for all efforts to answer questions from previous comments. I appreciated that authors added schematic depictions in current Figure 1, graphical methods in current Figure 4, and shuffle control in current Extended Data Figure 6-1 for reference. The remaining question I have below is related to my third question from previous comments. It would be great to learn whether the results found in both rate-coding and regression subspace models could be integrated and provide broad insights.

> We appreciate the reviewer's positive evaluation of our previous revision. Below, we reply to each comment raised by the reviewer.

Major comments:

1) Around 40% units in either cOFC (36.3%, Figure. 3C) and hippocampus (45%, Figure. 3H) populations do not selectively respond to task variables from conventional rate-coding analyses. Based on this, I would expect that there is no correlation when plotting regression coefficients (b1 and b2) included only non-significant units (like current Figure 3E and 3J, but without selective units). Therefore, I wonder whether or not authors have tried to recruit only selective units into the analysis of two-steps dynamical model? This concern is about uncorrelated data points may lead to biased eigenvectors and eigenvalues after applied PCA (A simple example is explained here: https://doi.org/10.1111/evo.13835). If this concern is reasonable, could authors take it into account and find a way to make corrections in the two-step dynamical model? This correction may take both rate-coding and dynamical models into account and improve the results in two experiments, especially in experiment 2. If this concern has no influence on the current results, please explain why.

> We thank the reviewers for the insightful comment regarding the potential influence of the non-significantly modulated neurons on PCA. We have carefully considered the implications raised in the paper “Be careful with your principal components”, and acknowledged the importance of addressing this issue in our analysis especially with respect to the following points.

First, we recognize that the aforementioned paper focused on the effect of unrelated trait on PCA, as shown in the figure 2 bottom in the paper. This is clearly important points for our analysis, because if we were to include task parameters that have minimal modulation on neural activity, we would encounter the same problems discussed in that paper. However, it is important to note that our experiments do not include such unrelated features or task parameters. In our analysis, all of the task parameters included have been shown to exert a significant influence on the activity of neuronal populations, as evidenced by the conventional rate-coding analyses shown in Fig. 3. Therefore, we can confidently assert that the concern raised by the reviewer (and so in the referenced paper) does not apply to our study.

Second, the reviewer inquired whether attempted to recruit only selective units in the analysis of two-steps dynamical model. We did not perform such an analysis for two main reasons. First, if we were to exclude activity without significant modulations, it would result in an inability to construct the regression matrix because there would be insufficient data in each time bin without significant modulation. Second, when using a narrow time bin such as the 20ms-bin, detection of statistical significance in each neural activity becomes less reliable due to the lower signal-to-noise ratio. Therefore, we decided to include all activity of neurons in this PCA analysis, which is a comprehensive representation of entire neuronal population.

We acknowledge the importance of the first point raised by the reviewer and have therefore added the following sentences to the Discussion section in our revision.

Discussion, P39

“We would like to caution researchers against using non-modulated task parameters in PCA, i.e., task parameters that show minimal neural modulation, as the PC’s derived in such cases were less biologically meaningful (Bjorklund, 2019).”

Minor comments:

1) To include the main feature, regression subspace, of your 2-steps dynamical model in the current Figure 1, could authors add a sample diagram as the second step in between neural modulation and PCA?

> Thank you for the reviewer's suggestion. We have tried to include the regression sub-space into the Figure 1 that makes the schema more precise. However, we did not find a suitable way to include and represent the regression subspace in this cartoon. Thus, we decided to include no sample diagram in the second step into this figure. We appreciate the reviewer's helpful comment, again.

2) To better understand the distributions of regression coefficients (b1 and b2) in Figure 3E and 3J, could authors color-label significant neurons by task variables (e.g. Figure 3C ad 3E) and count how many significant units in each quadrant?

> We appreciate the reviewer's important comment regarding the visualization of significant neural modulations in Figures 3E and 3J. In response to this suggestion, we have made the following changes to the Figures, although not exactly as suggested by the reviewer.

This is because Fig. 3E contains both regression coefficients but Fig. 3J contains only a part of item modulations.

In the revised Figure 3E, we have introduced color labels to represent a significant neural modulation based on either probability or magnitude. The plots are now color-coded, and we have counted the number of gray plots within each panel. Among the 190 cOFC neurons, we observed 18 (9.5%), 16 (8.4%), 40 (21.1%), 56 (29.5%), 53 (27.9%), 39 (20.5%), and 36 (18.9%) gray plots, respectively. It is worth comparing these values to sum of all three neural modulations shown in Fig. 3D.

For Figure 3J, we took a similar approach, coloring plots based on significant main effect of item. This is because we plotted the population activity for item

modulation of this figure. We then counted the number of gray plots, which were 18 (3.1%), 19 (3.2%), 44 (7.5%), 62 (10.5%), 47 (8.0%), 42 (7.1%), 40 (6.8%) among the 590 HPC neurons. It is also worth to compare these values with the green bars in Figure 3I. Note that the number of neurons counted is slightly different from that in Figure 3I. This occures because we plotted the regression coefficient excluding the interacting term, which is exactly the same matrix used in the dynamical analysis. This aspect of the conventional analysis serves as a link between our results and the dynamical analyses.

We believe that these modifications provide a clearer representation of the significant neural modulations in our study and improve the interpretability of the results.

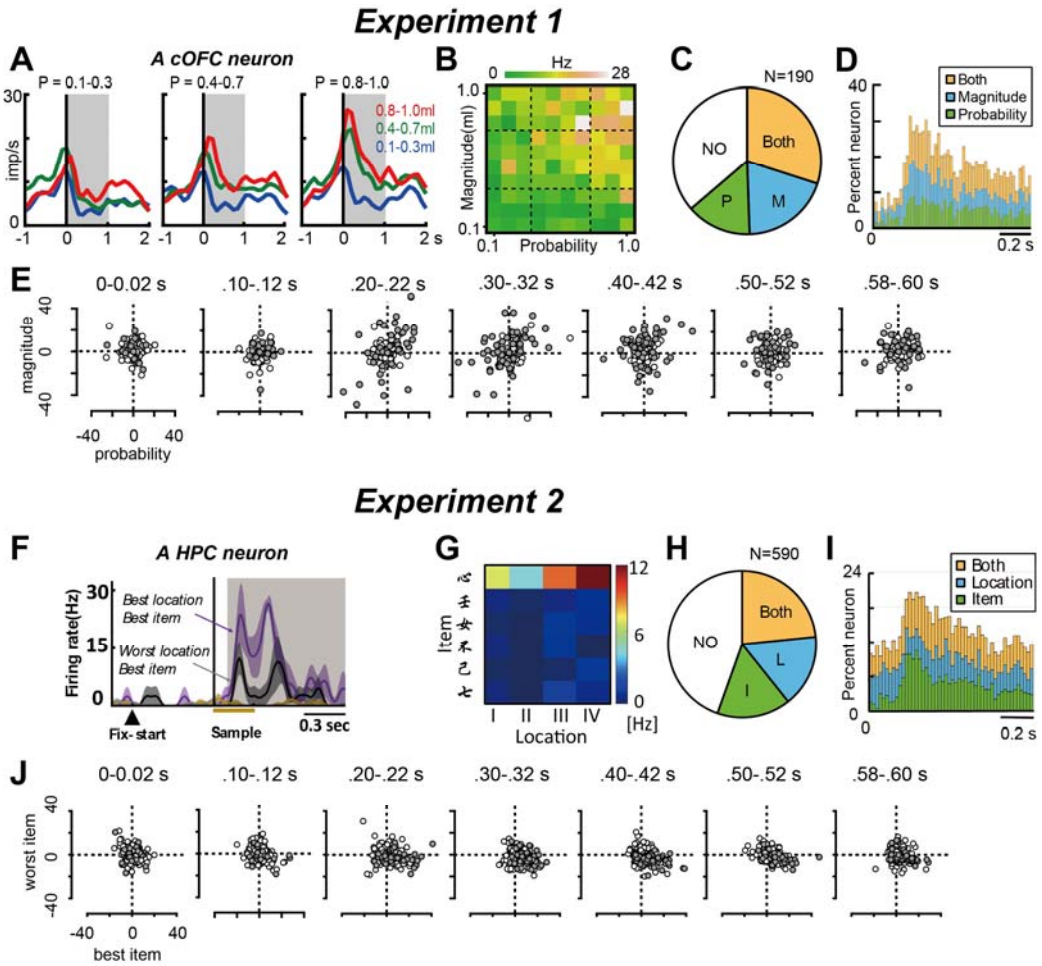

Figure 3. Example activity of neurons during the single-cue and ILR tasks.
(A) An example activity histogram of a cOFC neuron modulated by the probability and magnitude of rewards during the single-cue task. Activity aligned with cue

onset is represented for three different levels of probability (P, 0.1–0.3, 0.4–0.7, 0.8–1.0) and magnitude (M, 0.1–0.3 mL, 0.4–0.7 mL, 0.8–1.0 mL) of rewards. Gray hatched areas indicate the 1 s time window used to estimate the neural firing rates shown in B. Histograms smoothed using a Gaussian kernel (σ = 50 ms). (B) An activity plot of the cOFC neuron during the 1 s time window shown in A against the probability and magnitude of rewards. (C) The percentage of neural modulation types detected in 1 s time window shown in A; the probability (P), magnitude (M), both (Both), and non-modulated (NO). (D) Percentages of neural modulation type detected in the 0.02 s time bins during the 1.0 s after cue onset. The scale bar indicates 0.2 s. (E) Regression coefficient plots for the probability and magnitude of rewards estimated for all cOFC neurons in Exp. 1. Regression coefficients in the 0.02 s time bin shown every 0.1 s during the 0.6 s after cue onset (0–0.02 s, 0.10–0.12 s, 0.20–0.22 s, 0.30–0.32 s, 0.40–0.42 s, 0.50–0.52 s, and 0.58–0.60 s). Filled gray indicates significant regression coefficient for either probability or Magnitude at $P < 0.05$. (F) An example of an HPC neuron showing sample-triggered sample–location signals and item signals. A 0.08–1.0 s time window after sample onset was used to estimate the neural firing rates shown in G. Histograms are smoothed using a Gaussian kernel (σ = 20 ms). (G) An activity plot of the HPC neuron during the time window shown in F against item and location. (H) The percentage of neural modulation types detected in the 0.08–1.0 s window shown in F; item, location, both (Both), and non-modulated (NO). (I) Percentages of neural modulation types detected in the 0.02-s time bins during the 1.0 s after sample onset. (J) Regression coefficient plots for the best and worst items estimated for all HPC neurons in Exp. 2. Filled gray indicates significant regression coefficient for item at $P < 0.05$ using ANOVA without interaction term. The location modulation was not shown because we showed changes of neural modulation by the sample stimulus, whereas the location had already been provided to the monkeys. Figure 3A, B, and D were published in Yamada et al., 2021.

3) Typo in Figure 9: the title of exemplary plot in the bottom of Figure 9 should be PSTH.

> We corrected it and added the abbreviation in the legend for PSTH.