Stability and Predictability Code in Higher-Order Neuronal Correlations

Emili Balaguer-Ballester[1, 2], Ramon Nogueira[3], Juan Abolafia[4], Ruben Moreno-Bote[5, 6, 7], Mavi Sanchez-Vives[4, 8]

1. Department of Computing, Faculty of Science and Technology, Bournemouth University, Poole UK, **2**. Bernstein Center for Computational Neuroscience, Mannheim, Germany. **3**. Center for Theoretical Neuroscience, Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, USA.**4**. IDIBAPS, Barcelona, Spain, **5**. Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain, **6**. Center for Brain and Cognition, Merce Rodoreda building (Ciutadella campus), Barcelona, Spain, **7**. Serra Hunter Fellow Programme, Universitat Pompeu Fabra, Barcelona, Spain, **8**. ICREA, Barcelona, Spain.

Introduction

The functional role of the observed neural and behavioural variability in repetitions of the same task is a fundamental question in systems neuroscience [1-3]. However, the relationship between trial-by-trial shared variability (noise correlation) and behavioural performance is heterogeneous [4]. For instance, it has been proposed that neuronal pairwise correlations might not always serve as a proxy for behavioural performance, since only the variability along the encoding axis is detrimental to information transmission [5, 6].

In this study, we investigate the intricate relationship between the predictability of optimal choices, noise correlations, and stable states in rodent lateral orbitofrontal cortex (lOFC) ensembles. The OFC has been associated with multiple behaviourally relevant variables in the decision-making task space, such as outcomes expectations that guide action [7]. However, unlike in other frontal areas, the OFC signature of whether optimal choices are or are not predictable from previous trials outcomes is less established [7, 8].

Methods

We used a two-choice interval-discrimination task in which the previous trial outcome enables the animal to infer the optimal choice, but in which the upcoming stimulus is not always predictable (Fig 1a). In short, the animal must access the central socket to trigger a sequence of two pure tones (T1, T2) separated by either a short or a long inter-tone interval (ITI); by nose-poking either the left socket (for short ITI, light orange shades, Fig. 1a) or the right socket (for long ITI, darker shades) to retrieve the reward. After an incorrect trial, the previous ITI was repeated, and thus the upcoming trial is termed “predictable”. Otherwise, the ITI was randomly drawn (rendering “unpredictable” upcoming trials), which grade the task difficulty (Fig 1a, top). We identified four distinct trial outcomes, the correct choice (blue, Fig. 1b) and three different incorrect behavioural responses (red, grey, and green in Fig. 1b).

The dataset consisted of 82 neurons in ensembles of up to n=8 units from 3 Wiskar rats’ lateral orbitofrontal cortex, see details in [9]. Consider the $n \ x \ 1$ firing rates vector $\boldsymbol{x}(t,T)$, in which each component $x_i(t,T)$ is the i^{th} unit firing-rate at a time bin t in a trial of duration T; and $< x_i(t) >$ is the mean over the entire trial duration. The θ^{th}- order Pearson correlation coefficient for an n-units ensemble is defined as:

$$Corr(x_1, \ldots, x_n; \theta, t) = \frac{\sum_T (x_1(t,T) - < x_1(t) >)^{m_1} \cdots (x_n(t,T) - < x_n(t) >)^{m_n}}{\sqrt[\theta]{(\sum_T (x_1(t,T) - < x_1(t) >)^{\theta})^{m_1} \cdots (\sum_T (x_n(t,T) - < x_n(t) >)^{\theta})^{m_n}}}, \qquad (1)$$

where $\sum_{i=1}^{n} m_i = \theta, \ m_i \in 0, 1, \ldots, \theta - 1; \ \theta > 1$, and trial-averages were specific to each response category. Thus, we loosely refer to eq. 1 as "noise decision correlations" (hereafter "correlations"); since all trials share the same outcome (Fig 1c). The robust calculation of all possible θ^{th}- order correlations (eq. 1) is demanding in datasets size and computational cost. However, eq. 1 can be alternatively derived from a specific $\mathcal{D}$- dimensional reproducing kernel-Hilbert space $\mathcal{H}^{\mathcal{D}(\theta)}$, where the inner product in this space is the well-known θ^{th}- order inhomogeneous multinomial kernel [10]. The j^{th} component of a $\mathcal{H}^{\mathcal{D}(\theta)}$ vector $\boldsymbol{\phi}$ (of dimension $\mathcal{D} = \binom{n+\theta}{n} - 1$) is:

$$\phi_j(\theta, \boldsymbol{x}(t,T)) = \sqrt{\binom{\theta}{i_0, \ldots, i_n}} \cdot x_1(t,T)^{i_1} \cdots x_n(t,T)^{i_n}, \ j = j(i_0, .., i_n), \ i_k \in N^+, \qquad (2)$$

where $0 \leq i_0 < \theta, \ 0 \leq i_{k \neq 0} \leq \theta, \ \sum_{k=0}^{n} i_k = \theta$. The full expression of eq. (1) in terms of $\mathcal{H}^{\mathcal{D}(\theta)}$ for any θ is shown in [9]. For example, the conventional noise correlation $(\theta = 2)$ for any pair of distinct units x_{a_1}, x_{a_2}, is, straightforwardly, $\mathrm{Corr}(x_{a_1}, x_{a_2}; \theta = 2, \ t) = \sum_T \tilde{\phi}_j(\theta = 2, \tilde{\boldsymbol{x}}(t,T))$, where $j = j(i_{a_1} = 1, i_{a_2} = 1)$, $a_1, \ a_2 > 0$ is a component of the vector (eq. 2) normalized by the binomial coefficient, and $\tilde{\boldsymbol{x}}(t,T)$ are the z-scored ensemble rates. The correspondence between eq. 1 and $\mathcal{H}^{\mathcal{D}(\theta)}$ dimensions (eq. 2) enabled us to explore the full high-dimensional space of all possible θ^{th}- order correlations (eq. 1), not directly accessible computationally, leveraging kernel or gaussian process classifiers [9].

Here we used a basic bayes-optimal classifier, a kernelized discriminant operating in $\mathcal{H}^{\mathcal{D}(\theta)}$, like in [10]. We reported the maximum-a-posteriori estimation of the causally (in future trials) 6-fold cross-validated decoding error (DE %) for the distinct choice outcomes; and the trajectory divergence index (DT %), a measure of the degree of attracting behaviour for each trial T_0 trajectory $\boldsymbol{x}(t, T_0)$ (see details in e.g., [9, 11, 12]). Kruskall-Wallis and Wilcoxon rank sum test were used (Fig. 1d-g, *p<0.01; **p<0.001, Bonferroni corrected). Orange boxplots in Fig. 1c-e refer to n=1000 bootstraps in which the trial outcome was shuffled before decoding.

Results highlights and concluding remarks

Stability of predictable choices. We utilized this methodology to assess the effect of neuronal correlations in the decodability of choice outcome. Fig. 1b shows an example of a reduced state space representation derived from a kernelized discriminant (DC=discriminant coordinate) operating in $\mathcal{H}^{\mathcal{D}(\theta=3)}$, cross-validated in blocks of 40 future trials for a single ensemble. Only states representing correct choice outcomes remained stable for over 100 trials of the task (Fig. 1b, blue) and thus were effectively decoded (Fig. 1c left, DE) and showed attracting properties (Fig. 1c right, DT). In contrast, incorrect choice states randomly wandered in the space (Fig. 1b, red, grey, green), and hence they could not be decoded (Fig 1c, above bootstraps). Overall, when all data was considered, the correct choice (blue, Fig 1d) was optimally decoded versus all three incorrect choices (mean shown in black) in this state-space spanned by up to three-way interactions. Moreover, correct choice decoding was suboptimal for other $\mathcal{H}^{\mathcal{D}(\theta\neq3)}$ (now shown).

When the decoding analysis was restricted to trials following an incorrect choice (predictable trials), the same trend was observed, but significantly enhanced (Fig. 1e, right). However, and intriguingly, unpredictable correct choices could not be decoded any longer (Fig 1e, left), and the correct choice became indistinguishable from the other incorrect choice states. Thus, stability in the state space acted as a proxy for outcome predictability.

Trial-by-trial dynamics of noise correlations. The mapping between decoding in $\mathcal{H}^{\mathcal{D}(\theta)}$ and eq. 1 enabled us to study correlation changes during the trial. Positive $\theta = 3$ correlations were dominant and stronger for correct choices (example in Fig. 1f), consistently with the higher stability of this state (Fig. 1e, right) and with results in [4] for pairwise correlations; but here they were similar for predictable/unpredictable trials (not shown), and thus could not fully explain Fig. 1e results.

However, the less frequent (~25%) triplet-wise negative correlations had a more interesting pattern. For unpredictable trials, they also discerned between correct and incorrect choice outcomes (Fig 1g, left). This counter-balancing effect of negative correlations in the stability of the correct-choice state is consistent with the high DEs shown in Fig 1e left. In stark contrast, for predictable trials, negative $\theta = 3$ correlations did not distinguish among choice outcomes (Fig 1g, right). In these trials, infrequent negative correlations were too weak to destabilize the correct-choice network state, explaining the lower DE for predictable correct choices in Fig. 1e (right). These effects were non-significant for pairwise and θ>3 correlations (not shown). Interestingly, negative correlations for predictable correct choices were very low before stimulus presentation (Fig. 1g, right, blue line before 200 ms), when the information needed for the optimal decision is already available. Thus, the hampering effect of negative correlations in decoding is negligible for predictable trials.

All in all, our results suggest that the successful processing of the task by small lOFC ensembles could map to long-lasting metastable states. Such metastable states gain stability when the optimal choice is deterministic and behaviourally relevant by attenuating negative triplet-wise correlations before stimulus presentation; and destabilize otherwise [9].

References

1. Renart A, Machens C. Variability in neural activity and behavior. Curr Op Neurobiol 2014; 25:211–220.
2. Nogueira R, Lawrie S, Moreno-Bote R. Neuronal Variability as a Proxy for Network State. Trends in Neurosci 2018; 41(4):170–173.
3. Balaguer-Ballester E. Cortical Variability and Challenges for Modeling Approaches. Front in Sys Neurosc 2017; 11.
4. Valente M, Pica G, Bondanelli, G et al. Correlations enhance the behavioral readout of neural population activity in association cortex. Nat Neurosci 2021. https://doi.org/10.1038/s41593-021-00845-1.
5. Moreno-Bote R, Beck J, Kanitscheider I, Pitkow X, Latham P, Pouget A. Information-limiting correlations. Nat Neurosci 2014; 17(10):1410–1417.
6. Kafashan M, Jaffe AW, Chettih SN. et al. Scaling of sensory information in large neural populations shows signatures of information-limiting correlations. Nat Commun 2021; 12, 473.
7. Nogueira R, Abolafia JM, Drugowitsch J, Balaguer-Ballester E, Sanchez-Vives M, Moreno-Bote R. Lateral orbitofrontal cortex anticipates choices and integrates prior with current information. Nat Commun 2017; 8:14823.
8. Namboodiri VMK, et al. Single-cell activity tracking reveals that orbitofrontal neurons acquire and maintain a long-term memory to guide behavioral adaptation. Nat Neurosci 2019; 22(7):1110–1121.
9. Balaguer-Ballester E, Nogueira R, Abofalia JM, Moreno-Bote R, Sanchez-Vives MV. Representation of foreseeable choice outcomes in orbitofrontal cortex triplet-wise interactions. PLoS Comput Biol 2020; 16(6): e1007862.
10. Balaguer-Ballester E & Lapish CC (shared 1st authorship), Seamans JK, Durstewitz D. Attracting dynamics of frontal cortex ensembles during memory guided decision making. PLoS Comput Biol 2011; 7:e1002057.
11. Balaguer-Ballester E, Tabas-Diaz A, Budka M. Can We Identify Non-Stationary Dynamics of Trial-to-Trial Variability? PLoS ONE. 2014; 9(4):1–13.
12. Lapish CC & Balaguer-Ballester E (shared 1st authorship), Seamans JK, Phillips AG, Durstewitz D. Amphetamine Exerts Dose-Dependent Changes in Prefrontal Cortex Attractor Dynamics during Working Memory. J Neurosci 2015; 35(28):10172–10187.

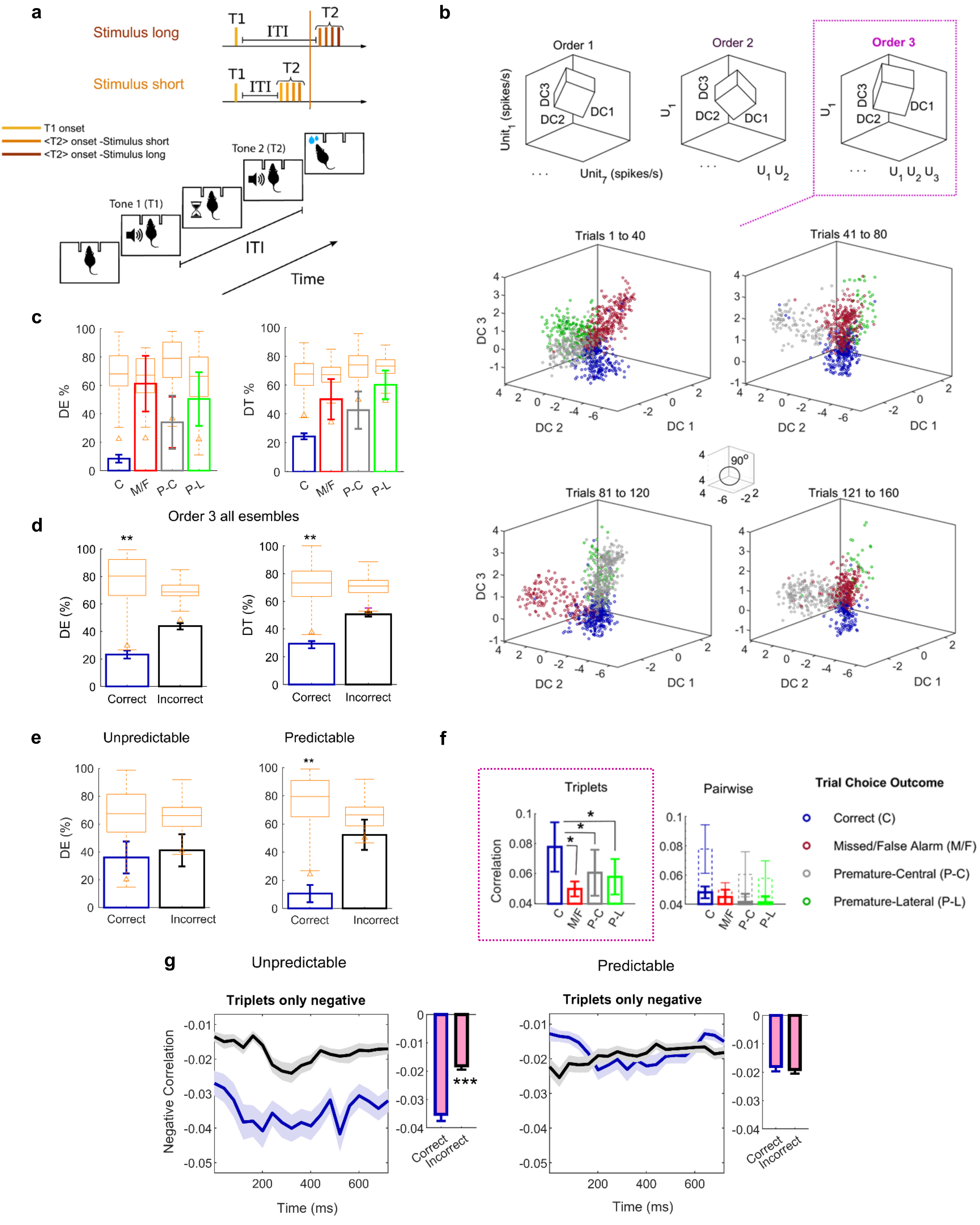

Figure 1. a. Experimental setting. **b**. Example of a reduced representation for a single ensemble via a kernelized discriminant ($\theta = 3$). **c.** Cross-validated decoding error (left) and trajectory divergence (right) for a single ensemble**. d.** Results for all data (black line: average for all incorrect choices). **e**. Decoding error for predictable/unpredictable trials. **f.** Positive noise correlations for a single ensemble (**b**). **g.** Mean negative triplet-wise correlations for all data. Details in [9].