Inferring readout of Distributed population codes without massively parallel recordings

Kaushik J. Lakshminarasimhan, Dora E. Angelaki, Xaq Pitkow

Summary

Information about task-relevant variables is often distributed among neurons across multiple cortical areas [1]. Neuronal responses are rarely independent of each other, but are correlated to some degree due to common input as well as recurrent message-passing. Consequently, determining how these neurons collectively drive behavioural changes requires not only examining how individual neurons are correlated with behaviour, but also estimating the correlated variability among neurons [2]. Precisely estimating the structure of correlated variability requires massively parallel recordings, which remains very difficult with current technology. Fortunately, it has recently been shown that the expansion in neural representation from sensory periphery will lead to a predictable pattern of correlations that ultimately limits the information content in brain areas downstream [3]. We examined the implications of these so-called information-limiting correlations for the readout of distributed population codes in a simple discrimination task. Surprisingly we found that both the behavioural precision, as well as the correlation of individual neurons with behavioural choice (*choice correlation*) were determined largely by the relative magnitudes of neuronal weights in the different brain areas and not on their specific pattern. We also found that, in the presence of information-limiting correlations, the choice correlations of neurons within an area should all scale by the same factor following inactivation of other potentially task-relevant brain areas. Together, our results lead to a novel framework for inferring how different brain areas contribute to behavioural response. Specifically, we show that the contribution of a brain area can be inferred simply by observing how the magnitude of choice correlations of individual neurons within the area and the behavioural precision are affected by inactivating other areas, thus obviating the need for large-scale recordings.

Supporting Materials

Background: Owing to redundancies generated by information-limiting noise, a broad range of decoding weights would all yield stimulus estimates that are nearly optimal. Consequently in large neural populations, the precise pattern of weights does not matter much and choice correlations C_k of single neurons approach ϑ/ϑ_k where ϑ is the behavioural threshold and ϑ_k is the threshold of neuron *k* [4]. We have shown previously that when limited information is distributed across multiple neural populations, the resulting choice correlations of neurons in population *z* are given by $C_{kz} = \beta_z(\vartheta/\vartheta_{kz})$ where β_z is once again independent of the precise pattern of weights, but depends on the overall magnitude of neuronal weights v_z in area *z* [5]. This theory is consistent with neural data in macaques [6]. Here we build on these results to develop a novel experimental framework for inferring the magnitude of weights v_z in different brain areas.

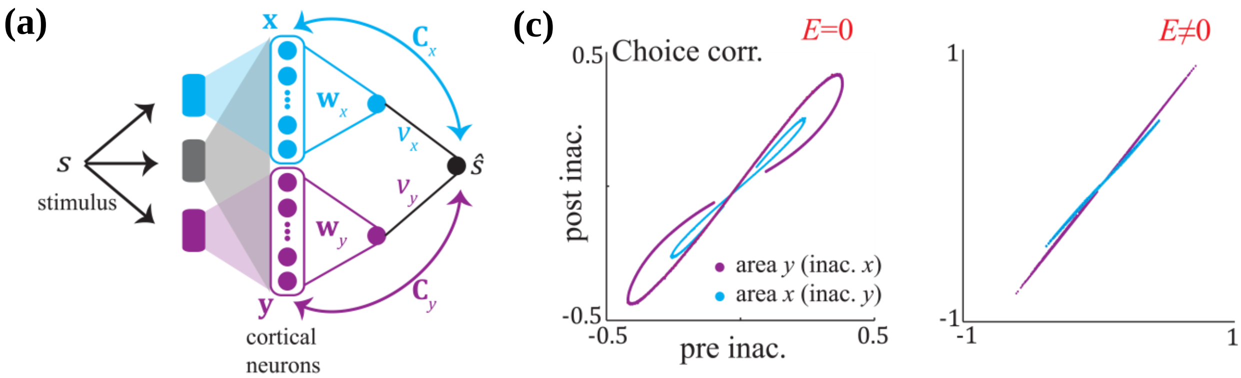
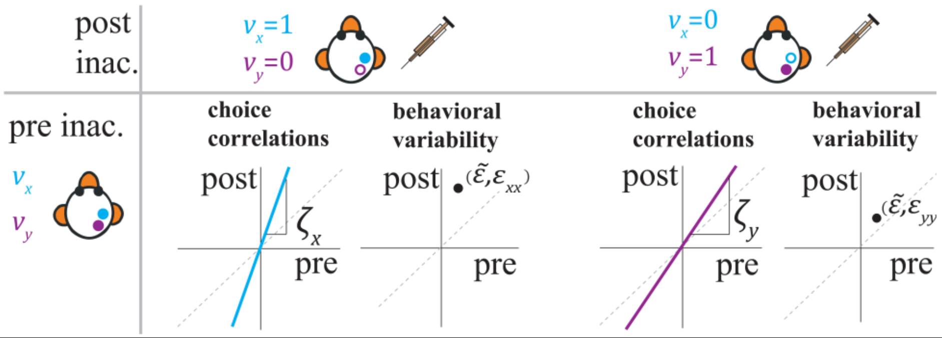


Figure 1. Information-limiting noise restricts pattern of choice correlations. (a) Feedforward network in which two cortical areas (*x* and *y*) receive information about stimulus *s* from independent as well as overlapping upstream sources. The areas are linearly readout with weights v_x and v_y to generate estimate $\hat{s}$ of the stimulus. **(b)** Covariance matrix is comprised of limited-range correlations (Σ_0) that do not limit information, as well as a small amount of information-limiting noise. **(c)** In the model that lacks information-limiting noise ($E = 0$), both magnitude and pattern of choice correlations in either area are affected following inactivation of the other area, whereas only their magnitude is altered in the model that incorporates it.

Results: We considered two populations of N neurons with mean activities $\mathbf{x}$ and $\mathbf{y}$ and sensitivities $\mathbf{x}' = d\mathbf{x}/ds$ and $\mathbf{y}' = d\mathbf{y}/ds$ in response to a particular stimulus s (**Fig. 1a**). Response covariance was defined as $\Sigma = \Sigma_0 + UEU^T$ where Σ_0 is covariance that does not limit information, and UEU^T corresponds to information-limiting noise where $U = \begin{pmatrix} \mathbf{x}' & \mathbf{0} \\ \mathbf{0} & \mathbf{y}' \end{pmatrix}$ is an $N \times 2$ matrix whose columns correspond to the directions of noise fluctuations and $E = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} \\ \varepsilon_{xy} & \varepsilon_{yy} \end{pmatrix}$ represents their covariance (**Fig. 1b**). Linear readout of the populations produces an unbiased estimate whose mean is given by $\langle \hat{s} \rangle = v_x \mathbf{w}_x^T \mathbf{x} + v_y \mathbf{w}_y^T \mathbf{y}$ where v_x and v_y denote the relative magnitudes on the corresponding unbiased weight patterns $\mathbf{w}_x$ and $\mathbf{w}_y$ such that $v_x + v_y = 1$. In this framework, the effect of inactivating area *z* (either *x* or *y* in this case) can be examined by setting $v_z = 0$ so that neurons in area *z* have no contribution to behavioural outcome. Inactivation of brain areas alters choice correlations $\mathbf{C}_z$ of neurons in active ones in a predictable manner, and this can potentially be exploited to infer the magnitude of weights in the different brain areas. However theories of population coding that ignore information-limiting correlations (i.e. $E = 0$) predict that inactivation of one area will affect both the *magnitude* and *pattern* of choice correlations in the remaining active brain area (**Fig. 1c** - left). Inferring weights from such results requires large-scale simultaneous recordings to estimate the precise structure of noise correlations in addition to choice correlations which can be obtained by recording only one neuron at a time [2]. In contrast, in the presence of information-limiting noise ($E \neq 0$), the altered choice correlations $\tilde{\mathbf{C}}_z$ of all neurons in area *z* differ from their original values $\mathbf{C}_z$ by the same scalar multiple (**Fig. 1c** - right). This is because the *pattern* of $\mathbf{C}_z$ depends mostly on the information-limiting mode $\mathbf{z}'$ with the fine structure of noise being largely irrelevant. Inactivation of other areas leaves the shape of this noise mode intact, and thus preserves the pattern of $\mathbf{C}_z$ while changing only its magnitude.

We used this result to develop a simple experimental paradigm for determining v_x and v_y using only single-cell recordings. Recording exclusively from single neurons in one area before and after inactivation of the other in behaving animals will yield all the measurements required to infer the quantities of interest (**Fig. 2**). Specifically, estimates of choice correlations and behavioural thresholds can be used to solve the system of equations below to infer the values of v_x and v_y and thus understand how the brain weights information available in different areas.



$$\zeta_x = \frac{\tilde{\varepsilon}}{v_x \varepsilon_{xx} + v_y \varepsilon_{xy}}$$

$$\zeta_y = \frac{\tilde{\varepsilon}}{v_x \varepsilon_{xy} + v_y \varepsilon_{yy}}$$

$$\tilde{\varepsilon} = v_x^2 \varepsilon_{xx} + v_y^2 \varepsilon_{yy} + 2 v_x v_y \varepsilon_{xy}$$

$$v_x + v_y = 1$$

Figure 2. Experimental paradigm for inferring magnitudes of weights in the two areas. Brain areas *x* and *y* (shown here in cyan and purple) can be selectively inactivated by injecting a pharmacological agent while monitoring behavioural variability and estimating neuronal choice correlations ($\mathbf{C}_x$ and $\mathbf{C}_y$) from the other area. $\mathbf{C}_x$ and $\mathbf{C}_y$ change by a factor of ζ_x and ζ_y following inactivation of *y* and *x* respectively, while behavioural variability changes from $\tilde{\varepsilon}$ to either ε_{xx} or ε_{yy}. Estimates of ζ_x, ζ_y, $\tilde{\varepsilon}$, ε_{xx}, and ε_{yy} can be used to solve for the values of v_x and v_y using the system of equations to the right. Note that there are four equations but only three unknowns, so one only needs to measure either ζ_x or ζ_y to solve the system. Alternatively, one could measure both to test their solution for self-consistency.

Conclusion. Information-limiting correlations generate redundancies in the neural code. We have exploited these redundancies to develop a coarse-grained theory of linear population codes. Our theory leads to a simple framework for inferring how brain areas drive behaviour, using only single-cell recordings.

References

[1] Britten. *Annu. Rev. Neurosci.* **31**, 389-410 (2008).
[2] Haefner, Gerwinn, Macke, Bethge. *Nat. Neurosci.* **16,** 235–42 (2013).
[3] Moreno-Bote, Beck, Kanitscheider, Pitkow, Latham, Pouget. *Nat. Neurosci.* **17,** 1410–1417 (2014).
[4] Pitkow, Liu, Fetsch, Chen, Angelaki, DeAngelis, Pouget. *Cosyne Abstracts.* Salt Lake City USA (2013).
[5] Lakshminarasimhan, Pouget, Pitkow. *Cosyne Abstracts.* Salt Lake City USA (2014).
[6] Lakshminarasimhan, Liu, Gu, Klier, DeAngelis, Pitkow, Angelaki. *Cosyne Abstracts.* Salt Lake City USA (2014).