Percept and the single neuron

Adrien Wohrer & Christian K Machens

A newly derived link between neural readout weights and choice signals in perceptual decision-making provides a fresh perspective on how percepts are formed from the activity of sensory neurons.

Decades of research have taught us that our mental lives are reflected in the activity of neurons. Yet when we perceive an uncertain stimulus—was that a friend hustling by on the other side of the street?—how many neurons are really implicated in the formation of the mental percept? When deciding on whether to shout "hello" or not, do our brains exploit all the sensory information available at the neural level? A first answer to these questions was given about 15 years ago, with the surprising finding that the activity of single sensory neurons can show detectable traces of a visual percept[1]. When a monkey has to make a decision based on weak or nonexistent visual information, small fluctuations of single-neuron activity in the visual cortex are slightly predictive of the animal's final choice. However, this observation alone does not tell us whether such neurons actively participate in forming the animal's percept or are simply influenced by it. In this issue of *Nature Neuroscience*, Haefner *et al.*[2] report significant conceptual progress in resolving this long-standing puzzle.

The authors studied the 'standard model' of percept formation in the context of a two-alternative forced-choice task (**Fig. 1a**): a subject is presented with a set of stimulus values *s* with little informative power, such as a noisy set of horizontally moving dots, and must decide whether $s > 0$ (overall motion toward the right) or $s < 0$ (overall motion toward the left). Neurons in visual cortex fire whenever the physical stimulus is close to their preferred stimulus. The relative strength of their firing can be interpreted as a vote for their preferred stimulus. In the standard model, the animal's

Adrien Wohrer is at the Group for Neural Theory, INSERM U960, Ecole Normale Supérieure, Paris, France, and Christian K. Machens is at the Champalimaud Neuroscience Programme, Lisbon, Portugal.
e-mail: christian.machens@neuro.fchampalimaud.org

percept *P*, meaning its internal estimation of the stimulus *s*, is formed by weighting the firing rates r_k with 'readout' weights β_k and summing them across the population. The readout weights capture the nature of percept formation in the task: if a neuron's readout weight is zero, it does not contribute to the percept; if the weight is nonzero, it contributes, however weakly. Finally, the animal's binary choice in each trial is computed by asking whether $P > 0$ or $P < 0$.

To determine each neuron's contribution to the visual percept, we therefore need to figure out the readout weights. Even though we can only monitor the neural activity and the animal's choice, this goal seems reachable: if a neuron's readout weight is large, any trial-to-trial variability in its firing rate will influence the animal's decision. This influence can be measured by the choice probability (CP)^, the probability of correctly predicting the animal's choice in each trial based on the neuron's activity (**Fig. 1b**). Unfortunately, however, a large CP value does not automatically imply a large readout weight. Imagine a neuron that receives synaptic inputs from a second neuron with a large readout weight (and thus a large CP value). Through this input, the first neuron's activity will partly reflect that of the second, leading to a substantial CP value. For all we can tell, however, the first neuron's readout weight may very well be zero.

More generally, researchers have noted that if the activity of pairs of neurons covaries from trial to trial—that is, if neural firing rates are correlated—then it is nontrivial to discern the link between CP values and readout weights. Although several previous studies have explored this link for a few specific scenarios, using computer simulations, Haefner *et al.*[2] now provide the general solution, by deriving the link between CP and readout weights analytically. The authors' main result is the following: a neuron's choice probability depends on

the sum over all of its firing rate covariances with other neurons, C_{kj}, weighted by their respective readout weights, β_j:

$$CP_k - \frac{1}{2} = \alpha \sum_{j=1}^{n} C_{kj} \beta_j \qquad (1)$$

where α is a normalization factor. This formula (which is only approximate; the authors also supply an exact version) provides the long-sought relationship between choice probabilities and readout weights in the presence of noise correlations. It essentially relies on the matrix product $\mathbf{C\beta}$, which is simply the trial-to-trial covariance between neural activities r_k and percept *P* in the standard model. Although it has been intuitively clear that CP measures something along these lines, Haefner *et al.*[2] now give mathematical grounding to this intuition.

Obtaining the readout weights β_k may seem straightforward now: we need to solve equation (1) by inverting the covariance matrix, C_{kj}.^ Unfortunately, that is impossible: because experimenters record from only selected ensembles of neurons for a finite number of trials, the covariance matrix cannot be fully determined. It is known only through samples of its elements C_{kj}, and with finite accuracy. How, then, are we to concretely exploit the beauty of the CP formula? Haefner *et al.*[2] provide two insights into this question.

The first insight—probably the more useful in practice—exploits the concept of optimality. Ideally, the organism should choose the readout weights $\mathbf{\beta}$ optimally, to maximize the sensitivity, or signal-to-noise ratio (SNR), of percept *P*. This optimal choice, Fisher's linear discriminant, is well known from the statistical literature and relies on the inverse of covariance matrix $\mathbf{C}$. As a result, when the optimal readout weights $\mathbf{\beta}^{(\text{opt})}$ is used, matrix $\mathbf{C}$ is eliminated from equation (1). Better, the exact expressions of $\mathbf{\beta}^{(\text{opt})}$ and proportionality term α lead to the following prediction: $CP_k^{(\text{opt})} \propto \text{SNR}(r_k)/\text{SNR}(P^{(\text{opt})})$. In words, if the animal
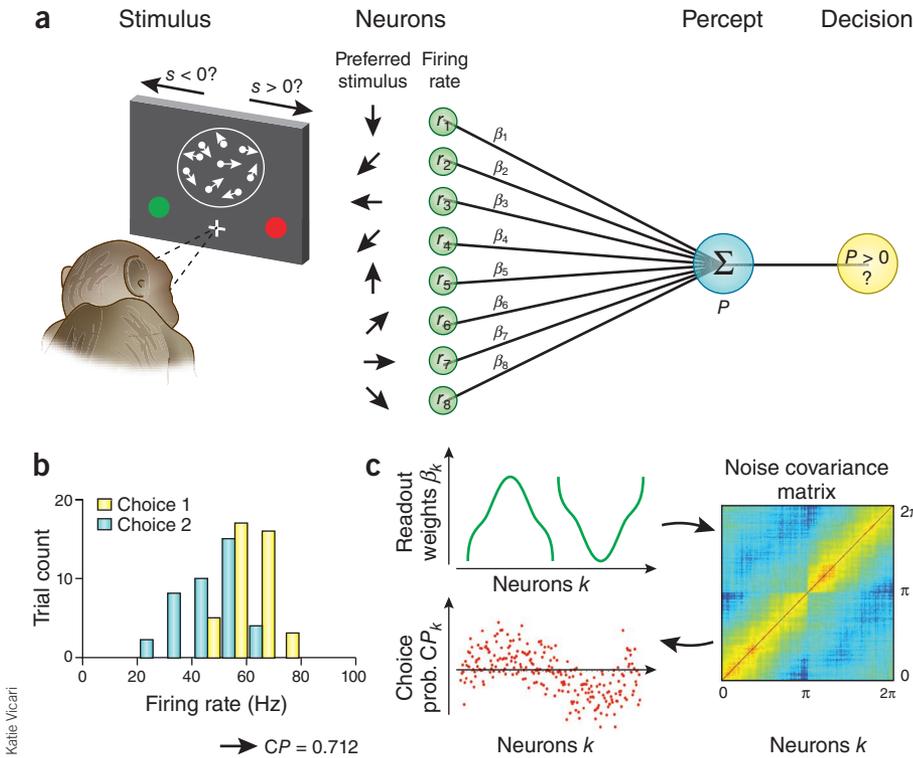
AU1

AU2

Figure 1 ^ The link between readout weights and choice probability signals in the standard model of percept formation. (**a**) The standard model of percept formation. Neurons with different tunings represent a stimulus (for example, the direction of a motion stimulus). The closer the stimulus is to its preferred tuning, the more strongly each neuron fires. ^ The resulting firing rates are weighted through the readout weights β_k and then summed to form the percept P. The sum is thresholded to give rise to the animal's decision. (**b**) Choice probability. If a neuron fired more strongly when the animal decided for one option (yellow, histogram across trials) than the other (blue), then the neuron's firing rate carries information about the animal's choice. The probability with which one can correctly guess the animal's choice on each trial is called the choice probability and is estimated from the overall amount of separation between the two distributions. Adapted from ref. 3 with permission. (**c**) Readout weights and noise covariance. The readout weights (top left) determine each neuron's contribution to the final percept. The choice probability of each neuron (bottom left) is given by a weighted sum over all readout weights. The summing weights, in turn, are determined by each neuron's noise covariances (right) with all other neurons. Adapted from ref. 2 with permission.

optimally extracts information from its sensory neurons, then each neuron's CP value will reflect the ratio between this neuron's sensitivity (its so-called neurometric threshold) and the overall sensitivity of the animal (its so-called psychometric threshold). Both types of sensitivity are routinely estimated from data, making this optimality test both simple and experimentally accessible. Of course, electrophysiological recordings may not pass this test. For example, by measuring CP values in a monkey faced with a novel behavioral task, Uka and DeAngelis found indirect evidence that the animal did not use all the neural information available, simply because it had not yet learned to do so[3]. Even in such a case, however, the test devised by Haefner *et al.*[2] allows us to quantify and assess any deviations from optimality.

The second insight—slightly more technical—sheds light on what information about

the readout weights β_k that we can, or cannot, reconstruct on the basis of experimental measures. The argument relies on the eigen decomposition of covariance matrix **C**. We know that **C** can always be decomposed into (at most) n orthogonal eigenmodes (or principal component axes) $\{\mathbf{v}_i\}_{i = 1...n}$. Each axis $\mathbf{v}_i$ defines one possible 'pattern of covariation' in the population, whose variance is captured by the corresponding eigenvalue λ_i. Then, if we write **C**P for the vector of the individual neurons' CP values, we can reexpress the CP profile as $\mathbf{C}P \propto \Sigma_{i = 1...n} \lambda_i v^{(i)} \mathbf{v}_i$, where $v^{(i)}$ is the projection of the readout vector **β** onto the ith principal component axis. According to this formula, any component $v^{(i)}$ of readout vector **β** corresponding to an eigenvalue $\lambda_i = 0$ for matrix **C** cannot affect CP values—and, thus, cannot be retrieved from the CP profile.

More generally, Haefner *et al.*[2] show how the difficulty of estimating the readout weights is

directly related to the amount of heterogeneity in the population. Experimenters recording from a small sample of the full population can only hope to build a 'smooth' estimate of the covariance matrix, based on other features of the population such as tuning. The fine structure of **C**, corresponding to heterogeneous and local patterns of covariance, will remain elusive—and so will all corresponding components of readout vector **β**. However, because any information extracted from these local structures will still affect CP values (generally making them smaller than their smooth predictions), it remains possible to experimentally estimate the amount of information extracted from the local structure of cortical connectivity.

The study of Haefner *et al.*[2] provides the first thorough interpretation of CP signals in the standard, linear model of decision-making. Although time is absent in their derivations, their main results carry over to temporal versions of the standard model, where the percept P is built from a linear sum of spike trains across neurons and across time. Thus, by looking at the temporal evolution of CP signals, experimenters may also be able to decipher when and how long sensory spikes are integrated to form the percept. Indeed, the authors' results are ripe for experimental testing, and such tests may clarify to what extent the standard model of percept formation is correct. For example, Nienborg and Cumming recently noted an oddity in the temporal evolution of CP signals in macaque secondary visual cortex, whose most parsimonious explanation would rely on the animal's choice feeding back into the sensory area and biasing its activity from trial to trial[4]. Whether this effect is indeed due to feedback, and whether it can be accounted for linearly, is unknown to this day. Thanks to the elegant analysis of CP signals of Haefner *et al.*[2], the main issues and conceptual problems have now been clarified. The ball is now in the court of experimental laboratories that can perform the suggested tests on their neural recordings and bring us another step closer to solving the puzzle of percept formation.

1. Britten, K.H., Newsome, W.T., Shadlen, M.N., Celebrini, S. & Movshon, J.A. *Vis. Neurosci.* **13**, 87–100 (1996).
2. Haefner, R.M., Gerwinn, S., Macke, J.H. & Bethge, M. *Nat. Neurosci.* **16**, xxx–yyy (2013). ^
3. Uka, T. & DeAngelis, G.C. *Neuron* **42**, 297–310 (2004).
4. Nienborg, H. & Cumming, B.G. *Nature* **459**, 89–92 (2009).