nature neuroscience

Inferring decoding strategies from choice probabilities in the presence of correlated variability

Ralf M Haefner[1–3], Sebastian Gerwinn[1,4,5], Jakob H Macke[2,4,6] & Matthias Bethge[1,2,4,7]

The activity of cortical neurons in sensory areas covaries with perceptual decisions, a relationship that is often quantified by choice probabilities. Although choice probabilities have been measured extensively, their interpretation has remained fraught with difficulty. We derive the mathematical relationship between choice probabilities, read-out weights and correlated variability in the standard neural decision-making model. Our solution allowed us to prove and generalize earlier observations on the basis of numerical simulations and to derive new predictions. Notably, our results indicate how the read-out weight profile, or decoding strategy, can be inferred from experimentally measurable quantities. Furthermore, we developed a test to decide whether the decoding weights of individual neurons are optimal for the task, even without knowing the underlying correlations. We confirmed the practicality of our approach using simulated data from a realistic population model. Thus, our findings provide a theoretical foundation for a growing body of experimental results on choice probabilities and correlations.

Understanding how external stimuli give rise to sensory percepts and how individual sensory neurons support this process are central questions of systems neuroscience. One of the crucial requirements for the claim that a particular group of neurons is critical for the generation of a perceptual event is that "fluctuations in the firing of some set of the candidate neurons to the repeated presentation of identical external stimuli should be predictive of the observer's judgment on individual stimulus presentations"[1]. Such correlations between the response fluctuations in a single neuron's firing rate and the subject's perceptual decision have been found in many cortical areas (for example, V1 (ref. 2), V2 (ref. 3), IT[4], MT[2,5–12], MST[13] and VIP[7]). They are usually quantified as choice probabilities[5]. The quantitative interpretation of choice probabilities has been problematic, however, as their connection to the read-out weight of a neuron is confounded by correlations among the sensory neurons[14]. For instance, a neuron that itself is not directly involved in a decision might display a significant choice probability purely because it is correlated with another sensory neuron that does directly contribute to the decision[15]. This indicates the importance of correlated variability and its structure, something that was recognized early on[14] and was highlighted again more recently[16]. A central challenge for all studies of choice probability is that this relationship between correlation structure and choice probabilities has not been characterized analytically. Thus, all previous studies are based on numerical simulations in which the key parameter, the correlation matrix, is very high dimensional—quadratic in the number of neurons in the considered population. This makes it infeasible to exhaustively explore the behavior of the system, fit it to empirical data, draw conclusions about the incompatibility of a particular model with a set of data, or acquire a deep understanding of the relationship between choice probabilities, sensory encoding and decision-making.

We mathematically derived the relationship between correlated variability, choice probabilities and read-out weights. We found simple relationships that explained how correlation structure and read-out weights together determine choice probabilities. This allowed us to prove earlier numerical results as well as a recent conjecture[16] about how choice probabilities depend on correlations for the previously much-studied case of a uniform read-out. Notably, we determined how our analytical solution allows one to infer aspects of the decoding weight profile from empirically observed neuronal correlations and choice probabilities. Finally, we derived a simple test for whether the read-out mechanism is optimal for the task, even in the absence of any knowledge about correlations.

RESULTS

Choice probabilities, correlations and read-out weights

We followed previous studies in modeling the decision-making circuit by assuming that a decision-making area linearly reads out and sums the activity of a population of neurons in a sensory area[14,16,17] (**Fig. 1**). A noise stimulus was presented and the task was to decide whether the stimulus contains more power for stimulus value s_1 or for s_2 (**Fig. 1a**). Examples may be a stimulus consisting of randomly moving dots, with the task to decide whether the net motion is in one or another direction[5,18], or a stereo image with the task to decide whether it is in front or behind the fixation point[3,19]. We examined a population response with each of n neurons in the population emitting r_k spikes in a given trial (**Fig. 1b**). These neural responses will, in general, be correlated, and we modeled this correlated variability by a multivariate Gaussian distribution with covariance matrix **C**,

[1]Max Planck Institute for Biological Cybernetics, Tübingen, Germany. [2]Bernstein Center for Computational Neuroscience, Tübingen, Germany. [3]Volen National Center for Complex Systems, Brandeis University, Waltham, Massachusetts, USA. [4]Werner Reichhardt Centre for Integrative Neuroscience, Tübingen, Germany. [5]Institute for Information Technology (OFFIS), R&D Division Transportation, Oldenburg, Germany. [6]Neural Computation and Behaviour Group, Max Planck Institute for Biological Cybernetics, Tübingen, Germany. [7]Institute of Theoretical Physics, University of Tübingen. Correspondence should be addressed to R.M.H. (ralf.haefner@gmail.com).

Received 3 October 2012; accepted 14 December 2012; published online 13 January 2013; doi:10.1038/nn.3309

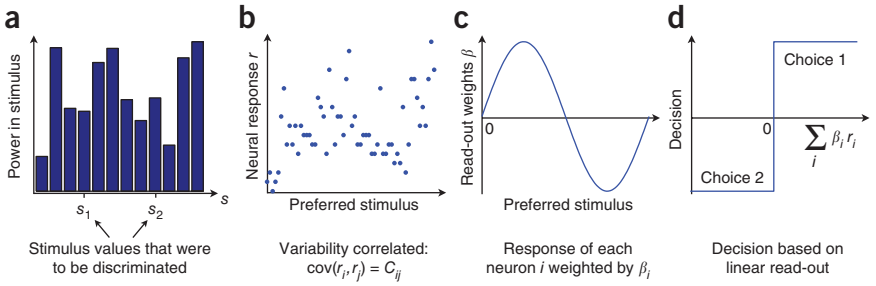

Figure 1 Illustration of model setup. (**a**) A noisy stimulus is presented containing power at two stimulus values s_1 and s_2. The task was to discriminate between s_1 and s_2. (**b**) Neuronal population response (neurons are sorted by preferred stimulus). The neural responses on each trial were variable and this variability was correlated (cov, covariance). (**c**) The activity of each neuron was weighted differently depending on their identity (here, preferred stimulus). (**d**) The decision was based on the sum of the weighted neural responses. Depending on the sign of the sum, either choice 1 or choice 2 was reported.

where C_{jk} is the covariance between the responses of neuron j and neuron k, and C_{kk} is the variance of neuron k. $\mathbf{C}$ is also called the noise covariance matrix and, when normalized by the variance, the noise correlation matrix. We modeled a linear read-out, which means that the activity of each neuron was multiplied by a weight β_k (**Fig. 1c**). The sign of the sum of the weighted responses determined the choice (**Fig. 1d**).

Consider the case of an ambiguous stimulus, one that contains equal evidence for either decision. Each sensory neuron will fire with a mean firing rate determined by its tuning function and a certain variability that we modeled as Gaussian distributed. If we split the trials into two groups, those that led to decision 1 and those that led to decision 2, and determine the distribution of firing rates, we may find that they are slightly different for each group. One is shifted to lower and the other to higher firing rates (**Supplementary Fig. 1a**). Such an effect has been observed in several sensory areas in cortex and its strength is usually quantified as choice probability[5,20]. More precisely, choice probability is defined as the probability that a random sample from the distribution for choice 1 trials is indeed larger than a random sample from the distribution for choice 2 trials. It is 0.5 if both distributions are identical and increases up to 1 as they are more and more separated. Although the stimulus will typically contain a small residual signal supporting one of the two decisions, this residual signal has a small effect on the differential firing distributions[5] and can be corrected for[19]. In the absence of any noise correlations, we could infer the read-out weight for a particular neuron from its choice probability alone: the larger the choice probability, the larger the neuron's read-out weight. However, in the presence of noise correlations, this is no longer true: a neuron might be assigned a zero read-out weight and show a large choice probability only because it is correlated with a neuron with nonzero read-out weight.

We derived the analytical relationship linking choice probabilities and read-out weights in the presence of arbitrary correlations (Online Methods):

$$CP_k = \frac{1}{2} + \frac{2}{\pi}\arctan\sqrt{2\xi_k^{-2} - 1}^{-1} \quad \text{with} \quad \xi_k = \frac{(\mathbf{C\beta})_k}{\sqrt{C_{kk}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{C\beta}}} \qquad (1)$$

CP_k is the choice probability of neuron k with respect to choice 1, which depends on the high-dimensional noise covariance matrix only through three numbers: C_{kk} (the response variance of neuron k), $(\mathbf{C\beta})_k = \sum_{j=1}^{n} C_{kj}\beta_j$ (the sum over all covariances of neuron k with all other neurons, weighted by their read-out weights plus the weighted variance of neuron k), $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{C\beta} = \sum_{k=1}^{n}\sum_{j=1}^{n}\beta_k C_{kj}\beta_j$ (the total variance summed across all neurons, weighted with the respective read-out weight plus the weighted sum over the covariances across all pairs).

A simplification of equation (1), allowing for an easier intuition, is given by its first order approximation

$$CP_k \approx \frac{1}{2} + \frac{\sqrt{2}}{\pi}\frac{(\mathbf{C\beta})_k}{\sqrt{C_{kk}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{C\beta}}} \qquad (2)$$

The error incurred by this approximation is very small, ranging from zero at $CP = 1/2$ to 5% for $CP = 1$ (**Supplementary Fig. 1b**). For clarity of exposition, we use this approximation to present our results for the choice probability in various scenarios.

Although our mathematical derivation relies on the assumptions of Gaussian response variability and perfect integration over the entire stimulus presentation, we confirmed its validity in more realistic settings (Poisson spiking, integration to bound[17], attractor-based decision-making[21] and pooling of noise only change choice probabilities by a scaling factor; see Discussion, **Supplementary Note** and **Supplementary Figs. 2–5**).

Interpretation of choice probabilities

The interpretation of choice probabilities, and particularly whether they tell us anything about how a neuron is read out by the decision-making mechanism, has long been a controversial topic[16,20]. On the one hand, it was observed that choice probabilities would only be large enough to be measurable in the presence of correlations in neuronal variability[14] and, based on simulations, that in large populations all neurons have the same choice probability regardless of their individual read-out weight[15]. On the other hand, empirical choice probabilities have been interpreted as information about how neurons are being read-out[12,21].

Equation (2) resolves this controversy by showing how choice probability is computed from a linear numerator, $\mathbf{C\beta}$, determining the shape of the choice probability profile, and a quadratic denominator $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{C\beta}$, scaling its magnitude. The linear part,

$$CP_k - \frac{1}{2} \propto \frac{1}{n}\sum_{j=1}^{n} C_{kj}\beta_j \qquad (3)$$

implies that, in addition to its own weight and its response variance, the choice probability of any neuron depends on its covariance with all other neurons in the population and their weights. Only one term in this sum depends on the read-out weight of the neuron itself; all of the other terms represent how the neuron is correlated with the other neurons in the population and what their read-out weight is. This means that, the larger the number of neurons contributing to the decision, the less the neuron's own read-out weight influences its choice probability. For the case of large correlated neuronal populations,

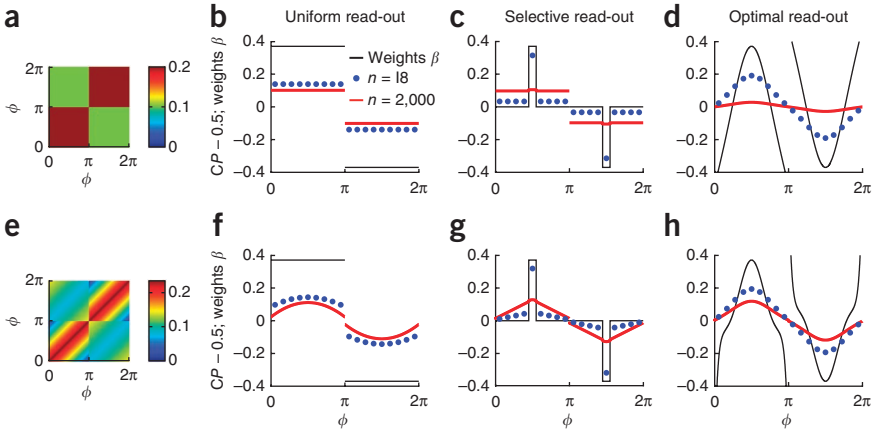

Figure 2 Choice probabilities for example cases. (**a**) Correlation matrix with a correlation coefficient of 0.2 between neurons in the same pool and 0.1 between neurons in different pools. (**b**–**d**) Choice probabilities produced by correlation matrix in **a** for different read-out profiles. The overall scale of the read-out weights (shown in black) was arbitrary (the decisions, and thus the choice probabilities, only depend on the sign of the summed responses weighted by the read-out weights, $\sum_{i=1}^{n}\beta_i r_i$, and are therefore unaffected if the weights β are all scaled by the same positive number). Blue dots represent choice probabilities implied by the respective read-out weight profile assuming there are a total of 18 neurons in the sensory population. Red curves represent choice probabilities produced assuming a population size of 2,000 neurons. (**b**) Uniform weights within each pool. (**c**) All weights were zero apart from the 5% of neurons, which were best-tuned to the stimuli that were distinguished in the task. (**d**) The weights were chosen to be linear-optimal for the task based on the assumed correlation structure (**a**). (**e**–**h**) Data are presented as in **a**–**d** assuming the correlation structure shown in **e**, which is based on empirical data[26]. We assumed homogenous populations with circular Gaussian tuning curves (von-Mises, $\kappa = 1.5$). The neurons were sorted by their preferred stimulus, ϕ. The task was to distinguish between stimuli moving in directions $\pi/2$ and $3\pi/2$.

the choice probability of a neuron essentially depends only on its covariance with the other neurons and on the other neurons' read-out weights. Notably, as each term containing a correlation is in fact a product of a correlation and a read-out weight, even in large neuronal populations choice probabilities reflect both correlations and read-out weights.

To examine our results, we used a classic coarse motion discrimination task[18] in which the subject has to decide whether the net motion in a random dot kinematogram is in one or the opposite direction. The neuronal population consists of n neurons with preferred directions ϕ between 0 and 2π. The two to-be-discriminated directions are $\pi/2$ and $3\pi/2$. This implies that neurons with preferred direction $0 < \phi < \pi$ support choice 1 and neurons with $\pi < \phi < 2\pi$ support choice 2. Note that this division into two pools is purely a result of the task context; the stimulus itself contains equal power in all directions. Our framework equally applies to fine discrimination tasks[11], tasks involving other stimulus dimensions[19,23] and other modalities[22].

We compared the expected choice probabilities for two different correlation structures and three different read-out weight profiles in a small and a large neuronal population (18 and 2,000 neurons, respectively; **Fig. 2**). For uniform correlations, choice probabilities contained a lot of information about the read-out weight of a neuron in small populations and very little in large ones (**Fig. 2c**). In practice, the choice probability profile for a selective read-out (**Fig. 2c**) is impossible to distinguish from that of a constant one (**Fig. 2b**), given typical measurement errors (s.e.m. of 0.05 for 100 trials). The majority of neurons with zero weights had a significant choice probability only because they were correlated with those neurons that had a nonzero weight, as noted previously[14,15]. For the case of optimal decoding (**Fig. 2d**), within each pool, neurons far away from the decision direction were generally subtracted from those whose preferred direction was close to the decision direction. This improves decision-making by subtracting (positively correlated) noise while leaving the signal largely unchanged[24,25]. The choice probabilities followed a roughly sinusoidal profile that reflects the shape of the neuronal tuning curves, as we show later, and the number of neurons in the population. We also examined a correlation structure that was based on empirical data[26] and its implications for the choice probabilities (**Fig. 2e**–**h**). Even if all neurons in a pool had the same weight, their choice probabilities could differ (**Fig. 2f**). If the correlations are largest between similar neurons, the neurons at the center of the pool have the largest choice probabilities because they are overall most correlated with the other neurons in their pool and least correlated with those in the other pool. Therefore, choice probabilities decreased in the direction of the pool boundaries, which was solely driven by the correlation structure, and not by their read-out weights (**Fig. 2g**). However, unlike for the case of uniform correlations, the choice probability profiles for uniform and selective weights remained different (**Fig. 2f**,**g**) even in arbitrarily large neuronal populations, making it possible to distinguish between these weight profiles by observing choice probabilities and correlations (**Supplementary Fig. 6**). The results for the optimal read-out are very similar for the two correlation structures shown here (**Fig. 2d**,**h**), apart from a substantial difference in overall magnitude for the large neuronal population.

The eigenvectors $\mathbf{v}^{(i)}$ (or principal components) of the noise covariance matrix, C, are those vectors for which multiplication with C changes their length, but not their direction: $C\mathbf{v}^{(i)} = \lambda_i \mathbf{v}^{(i)}$, where the scalar λ_i is called the corresponding eigenvalue. If the read-out weights β are equal to an eigenvector $\mathbf{v}^{(i)}$, then

$$\mathrm{CP} - \frac{1}{2} \propto \sqrt{\lambda_i}\mathbf{v}^{(i)} \tag{4}$$

as $C\boldsymbol{\beta} = \lambda_i \mathbf{v}^{(i)}$ and $\boldsymbol{\beta}^{\mathrm{T}} C\boldsymbol{\beta} = \lambda_i$ ($\mathbf{v}^{\mathrm{T}}\mathbf{v} = 1$ by definition). This implies that read-out weight profiles corresponding to eigenvectors with large eigenvalues will lead to large choice probabilities, and those profiles that resemble eigenvectors with small eigenvalues will lead to choice probabilities close to 0.5.

From now on we assume the correlation structure based on empirical data[26] (**Figs. 2e** and **3a**,**b**). For this correlation structure, the smaller the eigenvalue, the higher the frequency of the associated eigenvector (**Fig. 3c**,**d**). Because the actual read-out weights can be expressed as a linear combination of all the eigenvectors, $\beta = \sum_i v^{(i)}\mathbf{v}^{(i)}$, it is clear that, owing to measurement noise, empirical choice probabilities, contain more information about the low-frequency components of the read-out weight profile than the high-frequency ones (**Fig. 3e**). Consequently, the precision of the choice probability measurements imposes an upper limit on the frequencies for which the choice probabilities contain information, similar to how the Nyquist frequency is determined by the temporal resolution of a time series.

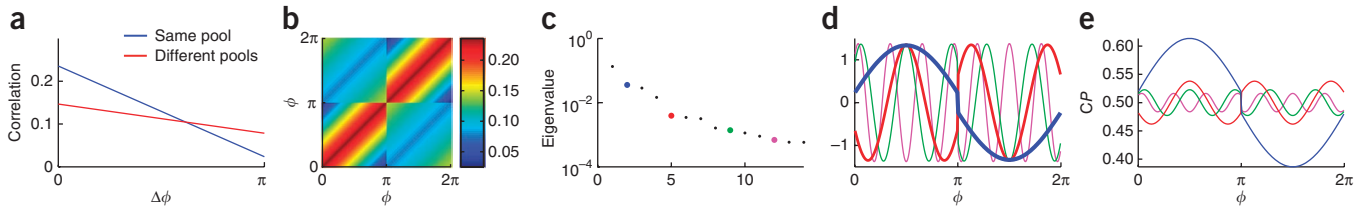

Figure 3 Correlation structure and its influence on choice probabilities. (**a**) Example correlation structure based on empirical data[26]. The dependence of the correlation between two neurons depending on their difference in preferred direction $\Delta\phi$ is shown for neurons belonging to the same pool (blue) and to different pools (red). (**b**) Correlation coefficient as a function of preferred direction as implied by a linear extrapolation of the profiles in **a**. (**c**) Largest 14 eigenvalues for the correlation structure in **b**. Colored eigenvalues correspond to eigenvectors shown in **d**. (**d**) Four of the eigenvectors belonging to the correlation matrix in **b** (see **Supplementary Fig. 8d** for the remaining eigenvectors). (**e**) Choice probabilities implied by read-out weights corresponding to the four eigenvectors shown in **d** for large neuronal populations.

As an example for the importance of this, we considered the case of uniform correlations within each of two large pools[15] (**Fig. 2a**). Such a correlation function has only two eigenfunctions with nonzero eigenvalues (**Supplementary Fig. 7b–d**), both of which are constant within each neuronal pool. This means that all deviations from a uniform read-out are present in a space about which the choice probabilities contain no information. This fact explains previous findings[15], based on simulations, that, regardless of the read-out weights of the individual neurons, the choice probabilities of different neurons in a pool are indistinguishable. In practice, however, a piecewise constant correlation function is unrealistic, and empirical correlation functions are likely to have a large number of nonzero eigenvalues (such as in **Fig. 3c**).

Reconstruction of read-out weights

If the complete correlation matrix for the sensory population as well as the choice probability of all neurons are known, then we can, in principle, invert equation (2) and deduce the read-out weights associated with individual neurons directly. However, there are two main challenges to performing this computation directly. First, in most cases, it is impossible to record from the complete neuronal population, and we can only record from a small subset of it. Second, measurement errors limit the accuracy of the reconstruction process. We overcame the first challenge by restricting ourselves to constructing a 'smooth model'. The central idea is to reconstruct the smooth read-out profile, $\beta(\xi)$, which is implied by the mean correlation function, $c(\xi_1, \xi_2)$, and mean choice probabilities, $CP(\xi)$. For our example case, we chose ξ to be the preferred stimulus direction, ϕ. However, other dimensions, such as neuronal sensitivity[27], neuronal type or alternative stimulus preferences[22], are possible. The second challenge can be addressed by restricting the reconstruction to the space spanned by those eigenfunctions whose eigenvalue is sufficiently large in relation to the measurement error in the choice probabilities. We obtained $v^{(i)}$, the projection of the read-out weights onto eigenfunction $\mathbf{v}^{(i)}(\phi)$, from the projection of the observed choice probabilities onto that same eigenfunction (Online Methods). Akin to representing a signal by its Fourier series in which all terms above the Nyquist frequency have been set to zero, we only considered the eigenfunctions with the largest eigenvalues.

We examined the feasibility of this approach using a realistic simulation of an experiment in which a small subset of a large neuronal population was recorded. We considered two scenarios. In the first, the true underlying read-out weights are constant within each pool, corresponding to a simple averaging of responses across each pool. In the second, the read-out is chosen optimally for each neuron to maximize discrimination performance for trials in which there is a signal in the stimulus (**Supplementary Note**). We now show that, using equation (18), it is possible to reconstruct the underlying weight profile from limited observations of correlations and choice probabilities and to reliably distinguish between the two example scenarios.

To account for neuronal heterogeneity, we assumed a true underlying correlation structure that was chosen randomly around the empirical mean[26] (shown in **Fig. 4a** for the 256 observed neurons; see Online Methods). **Figure 4b** shows the read-out weights for the two scenarios that we consider: constant weights and optimal weights. The dashed red line shows the optimal read-out weight profile for the corresponding average model, in which the variable correlation structure is replaced by its mean, and the variable neuronal tuning properties by their average quantities (assuming a homogenous population). The actual weights represented by red dots deviate from the optimal average read-out, as they also exploit small-scale differences between neighboring neuron's tuning properties and correlations to maximize performance. **Figure 4c** shows the measured choice probabilities for. The choice probabilities implied by the two weight profiles are systematically different for the two cases even in the presence of measurement noise (**Fig. 4c**), allowing us to differentiate between the two scenarios based on empirical data. An estimate of the full correlation structure from the empirical data was obtained by fitting a smooth correlation function to the noisy observed correlation coefficients (Online Methods and **Supplementary Fig. 8**). This allowed us to compute the eigenfunctions of C and to incorporate prior information about the read-out weight profile into the reconstruction. In our case, for instance, the rotational symmetry of the problem implies that only two of the seven eigenfunctions are relevant for the reconstruction (Online Methods). This leaves a two-dimensional space spanned by the eigenfunctions shown in **Figure 3e** in which we performed the reconstruction. The results of the reconstruction are shown in **Figure 4d**. The reconstructed weight profiles have to be compared with the ground truth (**Fig. 4b**) projected into the two-dimensional reconstruction space. Despite the fact that we only considered a two-dimensional space, the weight profiles preserve their characteristic properties: for the uniform weights scenario, all reconstructed weights have the same sign within each pool, and for the optimal weights scenario, the reconstructed weights have opposite signs near the decision directions and close to the pool boundary.

Restricting the weights reconstruction to the eigenfunctions with the largest eigenvalues has a noteworthy side-effect: the choice probabilities implied by the reconstructed model are larger than the choice probabilities implied by the full model. This is because the magnitude of the choice probabilities is proportional to the square root of the relevant eigenvalues, $\sqrt{\lambda}$, and the closer the actual read-out weight vector is aligned with eigenfunctions with small eigenvalues (which are ignored), the more the smooth reconstruction (which is based on large eigenvalues only) overestimates the choice probability magnitude (**Supplementary Note**). This effect is particularly large for non-smooth

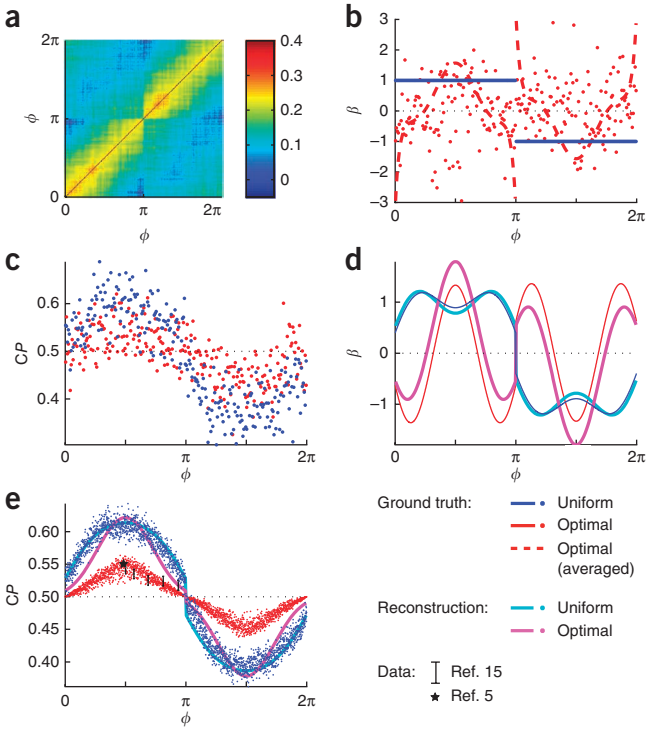

Figure 4 Reconstruction of weights from limited data in the case of a heterogenous neuronal population and limited data. (**a**) Ground truth correlation structure (sampled around the mean shown in **Fig. 3b**). (**b**) Ground truth weights β chosen to be constant in each subpopulation (blue) and linearly optimal (red). (**c**) Observed choice probabilities implied by correlation structure in **a**, weights in **b** and 200 trials. (**d**) Reconstructed weights (thick lines) based on the observed subpopulation of 128 neuron pairs. Thin lines represent the ground truth: the true weights as reconstructed from their projections onto the relevant eigenfunctions. (**e**) Ground truth choice probabilities for all neurons in our simulation (without any measurement noise, which is included in **c**). Cyan and magenta curves indicate choice probabilities profiles implied by the reconstructed smooth model (ignoring the heterogeneity in the neuronal response properties and in the correlation matrix). Black error bars indicate choice probability data from ref. 15 (s.e.m.). Black star indicates average choice probability found in ref. 5.

weight profiles that have significant power in higher frequencies. This applies to read-out profiles that are close to optimal, as they will have the highest power in eigenfunctions with small eigenvalues. The effect was apparent when we plotted the true choice probabilities alongside the choice probabilities implied by the smooth reconstructions (**Fig. 4e**). The average model recovered in the case of constant weights was very close to the simulated data. The choice probabilities implied by the average model for the optimal weights, on the other hand, were substantially larger than those observed. Notably, the shape of the choice probability profile was unchanged, and only the magnitude was greater. The heterogeneity in neuronal tuning curves is crucial here: if the underlying population were homogenous, the reconstructed model's choice probability predictions would be almost correct (data not shown), similar to the case of constant weights. This is because the weights profile optimal for a homogenous population has little power at higher frequencies that are ignored by the reconstruction.

We overlaid existing empirical[5,15] for comparison with our simulation results (**Fig. 4e**). We stress that the simulation results shown are not a fit to the choice probability data, but are the prediction from an optimal read-out based on a correlation structure extrapolated from the data in one of the two shown experiments[26] (red dots; **Fig. 4e**). To bring the magnitude of the choice probabilities for a uniform read-out into agreement with the data, one would need to invoke either a significant pooling noise more important than the entire sensory evidence or a decision mechanism that discards most of the sensory spikes (even in a reaction-time task like the one used to record this data). If the read-out was optimal, this magnitude problem would disappear or be at least greatly diminished. To distinguish between these two possibilities, one needs to consider the ratio of the reconstructed coefficients $v^{(1)}$ and $v^{(2)}$ as described below.

As each simulation is slightly different because of the simulated stochasticity in the experiment, we examined the distribution of reconstruction results for 1,000 independent simulations (**Fig. 5**). **Figure 5a** shows the average reconstructed profiles together with their variability across simulations. **Figure 5b** shows the distribution of the two reconstructed coefficients as two well-separated clouds. This demonstrates that an experimenter, who only has access to one point in this space, can distinguish between a uniform and an optimal read-out with high confidence.

Two potential complications may occur when trying to use $v^{(1)}$ and $v^{(2)}$ to decide whether a set of empirical data is compatible with a constant read-out or an optimal one: the subject may be using an internal integration-to-bound[17] or attractor-based[21] decision-making scheme rather than perfectly integrate the evidence over the entire trial, or noise may be added during the decision stage of the read-out[14]. Both possibilities have the same effect on the choice probabilities: a uniform scaling down by a factor compared with our analytical solution (**Supplementary Fig. 3**), leading to an equal uniform scaling of the $v^{(i)}$. To make our inference over the read-out weight profile independent of such factors, we considered the ratio of different $v^{(i)}$ values, here $v^{(1)}$ and $v^{(2)}$ (**Fig. 5c**). Although not as well separated as the coefficients themselves, their overlap was small and the power of this test was high (area under the receiver operating characteristic (ROC) curve is 0.93), even for the case of limited data that we considered (**Supplementary Fig. 9**). This indicates that, even in the absence of any information about pooling noise or for imperfect integration, it is possible to distinguish between uniform and optimal read-outs.

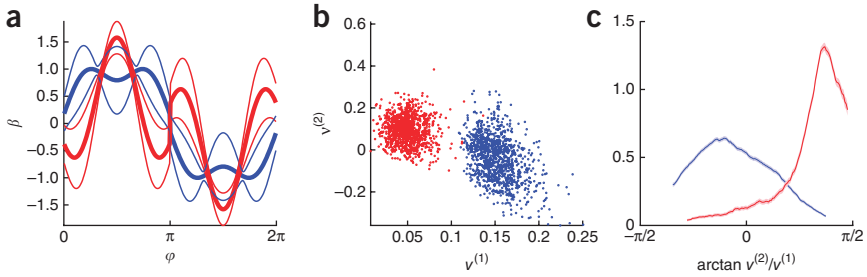

Figure 5 Reliability of reconstruction procedure for the simulated example case across 1,000 repetitions. (**a**) Reconstructed read-out profiles. Thick lines indicate mean reconstruction results and thin lines indicate s.d. Blue indicates the reconstruction for the uniform weight profile and red indicates the reconstruction for the optimal weights case. (**b**) Results for individual simulations in the reconstruction space spanned by the eigenfunctions with the two largest eigenvalues (and correct symmetry; see blue and red curves and dots in **Fig. 3c**,**d**). (**c**) Histogram of the ratios of the reconstructed projections onto the two relevant eigenfunctions (same as in **b**).

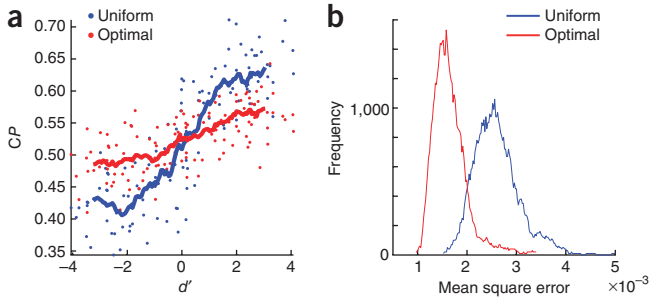

Figure 6 Optimality test. (**a**) Optimality test from equation (6) applied to the simulation in **Figure 4**. Blue and red dots represent the simulated data (constant and optimal weights, respectively, 1 dot per observed neuron). Thick curves represent running averages over 16 adjacent values. (**b**) Deviations from the proportionality relationship for the two cases as quantified by the mean square error.

Optimality test

We then used our framework to test the hypothesis that the weights are optimal for a particular task, even if the correlation structure is not known. This is possible because the formula that relates optimal weights to the correlation matrix,

$$\boldsymbol{\beta}^{\text{optimal}} \propto \mathbf{C}^{-1}[\mathbf{r}(s_1)-\mathbf{r}(s_2)] \tag{5}$$

is mathematically similar to the one relating choice probabilities to the correlation structure (equation (3)). Here, $\mathbf{r}(s_1)$ and $\mathbf{r}(s_2)$ are the neural responses to the two stimuli, s_1 and s_2, that are to be distinguished, and we assumed that C is approximately independent of the stimulus around the zero-signal stimulus[28] (**Supplementary Note**). Thus, if the read-out weights are optimal, the choice probability of a neuron is proportional to its sensitivity (d') in this task

$$CP_k - \frac{1}{2} \propto \frac{r_k(s_1)-r_k(s_2)}{\sqrt{C_{kk}}} \tag{6}$$

Note that, although the choice probabilities are measured on the basis of responses to zero-signal stimuli, the right-hand side of equation (6) requires knowledge of how the neurons' responses depend on the stimulus. However, to keep the subject motivated during a behavioral experiment, zero-signal trials are always interleaved with signal trials, and those responses can be used to compute $r_k(s_1) - r_k(s_2)$ (**Supplementary Information**).

We examined the relationship between the two sides of equation (6) (**Fig. 6a**). Although the running average of the simulated data for the case of constant weights showed a characteristic deviation from a straight line through the point (0;0.5), the running average for the simulated data in the optimal weights scenario lay close to a straight line. The scatter of the data points around the average was purely a result of measurement errors for CP_k, responses $r_k(s_1)$ and $r_k(s_2)$ and the response variability, C_{kk}, for each neuron: in the absence of measurement errors this relationship is perfect for an optimal linear read-out. By computing the average deviation of the empirically measured points from the best-fit straight line through (0;0.5) and comparing it with the deviations expected from measurement noise alone (for example, by resampling), one can statistically test whether the empirical data based on one experiment is compatible with an optimal read-out. We plotted histograms of the mean square error with respect to the best-fitting straight line for repeated simulated experiments (**Fig. 6b**). The small overlap between the histograms revealed the power of this test even for the limited amount of data used in our simulations (area under the ROC curve = 95%; for results based on more data from an electrode array, see **Supplementary Fig. 10**). What is particularly appealing about this relationship is that it involves only quantities that can easily be measured by traditional techniques (for example, single electrode recordings) and can therefore be applied to already existing data.

DISCUSSION

The discovery that the response of individual sensory neurons is correlated with an animal's behavior even when there is no signal in the stimulus[5] has been replicated for different decision procedures and in different sensory cortices[2–13] (for reviews, see refs. 1,16,17,20). In addition, appreciation of the crucial role of noise correlations for our understanding of sensory coding has grown[29,30]. Our results link these two lines of research by an analytical framework and show mathematically how choice probabilities, noise correlations and read-out weights are related. There are clear benefits of an analytical framework over numerical simulations: it allows us to fit parameters; to investigate much larger and more realistic populations, cutting computation time by several orders of magnitude; to extrapolate to arbitrarily large populations; and to devise rigorous tests for which types of models may or may not be compatible with the data.

As an application of our solution, and contrary to earlier conclusions[15], we found how one can reconstruct the decoding strategy of the brain for a classic perceptual decision task, under realistic conditions and with a reasonable amount of data using existing recording techniques. Although noise correlations were originally primarily seen as a complicating factor in inferring read-out weights from choice probabilities, our findings reveal that it is only because of their presence and their observed structure that we can use choice probabilities to infer anything at all about the read-out in large neuronal populations.

In addition, we used our framework to derive a test for whether the read-out of sensory neurons is optimal for a particular task with respect to tuning curves, response variability and noise correlations without requiring knowledge of noise correlations that have been difficult to measure. Applying this test to already existing data sets on a neuron-by-neuron basis should reveal whether decision neurons 'know' about tuning curves, response variability and noise correlations of their input neurons and account for them when pooling sensory information. This framework allows one, for the first time (to the best of our knowledge), to test the hypothesis of an optimal sensory read-out in common decision tasks.

At this point, there is little data available that combines measurement of interneuronal correlations with choice probabilities in a decision-making task[15,26,31]. However, with the increased use of population recording techniques[32–34], we expect more such data to become available very soon and our framework should prove readily applicable to infer decoding weights and shed light on the decoding strategies of the brain in various contexts.

Our analytical framework follows most previous studies in assuming a linear read-out[14,16,20,27]. In addition, we assume that the noise is additive and Gaussian and that the evidence is integrated for the entire stimulus presentation. In numerical simulations, we confirmed that our analytical results are excellent approximations if the spike counts are Poisson distributed (**Supplementary Fig. 2**). If the evidence integration was not perfect, but mediated by an integration-to-bound scheme[17] or attractor network[21], or degraded by pooling noise[14], the deviations from our analytical solution could be captured

by a single factor uniformly scaling down all choice probabilities (**Supplementary Figs. 3** and **4**). Any remaining deviations were substantially smaller than the measurement error resulting from high-quality recordings (500 trials).

A topic of recent debate has been the origin of choice probabilities[16,19]. We emphasize that, although our model appears to be explicitly bottom-up or feedforward, it is in fact largely agnostic about the source of its main ingredient, noise correlations. Although some noise correlation structures have traditionally been explained in a bottom-up way[35], correlations that depend on whether neurons belong to the same or to different pools are likely a result of top-down influences such as fluctuations in attention-like processes[26]. These sources of correlations are all compatible with our model assumptions. What is not covered by our assumptions is nonlinear dynamic feedback of the decision variable onto the sensory neurons. In general, given that our framework is necessarily an abstraction of more complicated mechanistic processes in the brain, our read-out weights are best interpreted as functional entities, representing the weights in an equivalent linear model that shows the observed responses and behavior (Online Methods and **Supplementary Fig. 5**).

Our results are not restricted to the coarse direction discrimination task, but equally apply to other domains (for example, disparity and vestibular direction[3,4,9,19,22,23], and fine discrimination[11]). It is unclear how our results translate to the interpretation of detect probabilities[7], as the symmetry in the two-choice task[21] was exploited in the derivation of our solution.

Given that, for most systems, it is only possible to observe a small subset of neurons, it is necessary to extrapolate the information for an observed subpopulation to the entire population. We showed how this problem can be solved by assuming correlations and read-out weights to depend on certain parameters, but not others; for example, preferred values for different stimulus parameters (such as, vestibular versus visual[22]), neuron type (such as, excitatory versus inhibitory), neuronal sensitivity[27] or brain area. Given current data limitations, we demonstrate such an extrapolation along only one dimension: the preferred direction of a neuron. In general, a limited number of observations means that weight variations above a certain frequency cannot be resolved. This frequency depends on the properties of the covariance structure and the quantity (number of neurons) and quality (number of trials) of the measurements. The use of electrode arrays will substantially increase the number of neurons and neuron pairs available, and thus the quality of the reconstruction. Furthermore, by applying our methods to multi-unit activity[15] or local field potentials[36], it may be possible to increase the number of trials by combining recordings across days, entailing a marked improvement in reconstruction accuracy (**Supplementary Figs. 9** and **10**).

Neuronal properties are known to be very heterogenous and we found that accounting for this variability was important for understanding empirical choice probabilities. We show that ignoring the variability in the read-out beyond what can be inferred from experimental data systematically overestimates the magnitude of the choice probabilities (**Fig. 4e** and **Supplementary Note**). In particular, read-out schemes that exploit neuronal heterogeneity to improve performance (as opposed to averaging) will have substantial power at higher frequencies, and thus lead to lower choice probabilities. The overall magnitude of choice probabilities may therefore provide valuable information as to the kind of read-out that the brain has implemented for any one task. Our work provides an example of the importance of accounting for neuronal heterogeneity in theoretical models. Simply averaging over variability in the data will lead to models that are not self-consistent. The inconsistency increases with the number of neurons, the variability in their tuning curve properties and in their response variabilities. The latter quantities are easily accessible empirically and, although already available in existing studies, are not routinely published on a cell-by-cell level. Our study adds to a growing body of literature that emphasizes the importance of reporting and modeling heterogeneity for understanding low-level function[37], sensory[38,39] and motor[40] circuits.

The right-hand side of equation (6) is sometimes called neuronal sensitivity and a positive correlation between choice probabilities and neuronal sensitivity has been observed in several studies[5,8,11,13,22]. It was alternatively attributed to increased correlations between the most sensitive neurons[14] or to a preferential read-out of the most informative neurons on the basis of intuition and numerical simulations[12,41]. We analytically derived the relationship between choice probabilities and neuronal sensitivity for the case of an optimal read-out code. Comparing predictions for a model with constant weights and those for a model with optimal weights, we found that the correlation between neuronal sensitivity and choice probability was positive in both cases. Notably, a positive correlation therefore cannot be taken as evidence that more sensitive neurons are weighted preferentially as has previously been assumed[41]. However, it is likely that an increased correlation between sensitivity and choice probabilities, given unchanged noise correlations (for example, over time[12]), is indicative of an improved read-out on the basis of sensitivities and correlations.

Equation (6) specifically predicts that the choice probabilities for an optimal read-out will gradually approach 0.5 as $r(s_1) - r(s_2)$ becomes zero, which happens at the decision boundary. This appears to be in contradiction with some empirical data[15] that shows a discontinuity at the decision bound with choice probabilities being significantly different from 0.5 even very close to the decision boundary on either side. A fine-discrimination study[11], on the other hand, found that choice probabilities did become 0.5 at the decision boundary, compatible with an optimal read-out in their task. Our prediction suggests that data collected for neurons with preferred direction halfway between the two choices is particularly informative in distinguishing between optimal and non-optimal read-out weights. By providing the exact relationship expected from an optimal read-out code, we provide the basis for a rigorous statistical test of whether decision neurons take into account response amplitude, response variability and correlations when reading out sensory neurons.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank A. Ecker and P. Berens for stimulating discussions and detailed comments on an earlier version of the manuscript, and H. Nienborg and B.G. Cumming for many helpful conversations. This work was partially supported by the German Ministry of Education, Science, Research and Technology through the Bernstein Award (FKZ 01GQ0601) (M.B.), the Bernstein Center for Computational Neuroscience (FKZ 01GQ1002), the German Excellency Initiative through the Centre for Integrative Neuroscience Tübingen (EXC307) and the European Commission (FP7-ICT-257005). R.M.H. acknowledges the hospitality of the Fiser laboratory at Brandeis University where this study was completed and financial support from the Swartz Foundation. Part of this research was done while J.H.M. was at the Gatsby Computational Neuroscience Unit, University College London.

AUTHOR CONTRIBUTIONS

R.M.H. conceived the research. R.M.H. and S.G. performed the analytical calculations and R.M.H. performed the simulations. All of the authors discussed the results. R.M.H. wrote the paper with contributions from the other authors. J.H.M. and M.B. advised at all stages.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nn.3309.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Parker, A.J. & Newsome, W. Sense and the single neuron: probing the physiology of perception. *Annu. Rev. Neurosci.* **21**, 227–277 (1998).
2. Grunewald, A., Bradley, D. & Andersen, R. Neural correlates of structure-from-motion perception in macaque V1 and MT. *J. Neurosci.* **22**, 6195–6207 (2002).
3. Nienborg, H. & Cumming, B. Macaque V2 neurons, but not V1 neurons, show choice-related activity. *J. Neurosci.* **26**, 9567–9578 (2006).
4. Uka, T., Tanabe, S., Watanabe, M. & Fujita, I. Neural correlates of fine depth discrimination in monkey inferior temporal cortex. *J. Neurosci.* **25**, 10796–10802 (2005).
5. Britten, K.H., Newsome, W., Shadlen, M., Celebrini, S. & Movshon, J. A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis. Neurosci.* **13**, 87–100 (1996).
6. Dodd, J.V., Krug, K., Cumming, B. & Parker, A. Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT. *J. Neurosci.* **21**, 4809–4821 (2001).
7. Cook, E.P. & Maunsell, J. Dynamics of neuronal responses in macaque MT and VIP during motion detection. *Nat. Neurosci.* **5**, 985–994 (2002).
8. Parker, A.J., Krug, K. & Cumming, B. Neuronal activity and its links with the perception of multi-stable figures. *Phil. Trans. R. Soc. Lond. B* **357**, 1053–1062 (2002).
9. Uka, T. & DeAngelis, G. Contribution of area MT to stereoscopic depth perception: choice-related response modulations reflect task strategy. *Neuron* **42**, 297–310 (2004).
10. Liu, J. & Newsome, W. Correlation between speed perception and neural activity in the middle temporal visual area. *J. Neurosci.* **25**, 711–722 (2005).
11. Purushothaman, G. & Bradley, D. Neural population code for fine perceptual decisions in area MT. *Nat. Neurosci.* **8**, 99–106 (2005).
12. Law, C.T. & Gold, J. Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nat. Neurosci.* **11**, 505–513 (2008).
13. Celebrini, S. & Newsome, W. Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. *J. Neurosci.* **14**, 4109–4124 (1994).
14. Shadlen, M.N., Britten, K., Newsome, W. & Movshon, J. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J. Neurosci.* **16**, 1486–1510 (1996).
15. Cohen, M.R. & Newsome, W. Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *J. Neurosci.* **29**, 6635–6648 (2009).
16. Nienborg, H. & Cumming, B. Correlations between the activity of sensory neurons and behavior: how much do they tell us about a neuron's causality? *Curr. Opin. Neurobiol.* **20**, 376–381 (2010).
17. Gold, J.I. & Shadlen, M. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
18. Newsome, W.T., Britten, K. & Movshon, J. Neuronal correlates of a perceptual decision. *Nature* **341**, 52–54 (1989).
19. Nienborg, H. & Cumming, B. Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature* **459**, 89–92 (2009).
20. Nienborg, H., Cohen, M. & Cumming, B.G. Decision-related activity in sensory neurons: correlations among neurons and with behavior. *Annu. Rev. Neurosci.* **35**, 463–483 (2012).
21. Wang, X.J. Decision making in recurrent neuronal circuits. *Neuron* **60**, 215–234 (2008).
22. Gu, Y., Angelaki, D. & DeAngelis, G. Neural correlates of multisensory cue integration in macaque MSTd. *Nat. Neurosci.* **11**, 1201–1210 (2008).
23. Nienborg, H. & Cumming, B. Psychophysically measured task strategy for disparity discrimination is reflected in V2 neurons. *Nat. Neurosci.* **10**, 1608–1614 (2007).
24. Abbott, L.F. & Dayan, P. The effect of correlated variability on the accuracy of a population code. *Neural Comput.* **11**, 91–101 (1999).
25. Chen, Y., Geisler, W. & Seidemann, E. Optimal decoding of correlated neural population responses in the primate visual cortex. *Nat. Neurosci.* **9**, 1412–1420 (2006).
26. Cohen, M.R. & Newsome, W. Context-dependent changes in functional circuitry in visual area MT. *Neuron* **60**, 162–173 (2008).
27. Law, C.T. & Gold, J. Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nat. Neurosci.* **12**, 655–663 (2009).
28. Bishop, C. *Pattern Recognition and Machine Learning* (Springer, New York, 2006).
29. Averbeck, B.B., Latham, P. & Pouget, A. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **7**, 358–366 (2006).
30. Cohen, M.R. & Kohn, A. Measuring and interpreting neuronal correlations. *Nat. Neurosci.* **14**, 811–819 (2011).
31. Gu, Y. *et al.* Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* **71**, 750–761 (2011).
32. Buzsáki, G. Large-scale recording of neuronal ensembles. *Nat. Neurosci.* **7**, 446–451 (2004).
33. Kerr, J.N. & Denk, W. Imaging *in vivo:* watching the brain in action. *Nat. Rev. Neurosci.* **9**, 195–205 (2008).
34. Stevenson, I.H. & Kording, K. How advances in neural recording affect data analysis. *Nat. Neurosci.* **14**, 139–142 (2011).
35. Shadlen, M.N. & Newsome, W. The variable discharge of cortical neurons: implications for connectivity, computation and information coding. *J. Neurosci.* **18**, 3870–3896 (1998).
36. Liu, J. & Newsome, W.T. Local field potential in cortical area MT: stimulus tuning and behavioral correlations. *J. Neurosci.* **26**, 7779–7790 (2006).
37. Marder, E. Variability, compensation, and modulation in neurons and circuits. *Proc. Natl. Acad. Sci. USA* **108** (suppl. 3): 15542–15548 (2011).
38. Padmanabhan, K. & Urban, N. Intrinsic biophysical diversity decorrelates neuronal firing while increasing information content. *Nat. Neurosci.* **13**, 1276–1282 (2010).
39. Ecker, A.S., Berens, P., Tolias, A. & Bethge, M. The effect of noise correlations in populations of diversely tuned neurons. *J. Neurosci.* **31**, 14272–14283 (2011).
40. Churchland, M.M. & Shenoy, K. Temporal complexity and heterogeneity of single-neuron activity in premotor and motor cortex. *J. Neurophysiol.* **97**, 4235–4257 (2007).
41. Jazayeri, M. Probabilistic sensory recoding. *Curr. Opin. Neurobiol.* **18**, 431–437 (2008).

ONLINE METHODS

Notation and derivation. We assumed that the decision is based on a linear combination of the responses $\mathbf{r} = (r_1,..,r_n)$ of a population of n sensory neurons

$$D = \sum_{k=1}^{n} \beta_k r_k \equiv \boldsymbol{\beta}^{\mathrm{T}}\mathbf{r} \tag{7}$$

where $\boldsymbol{\beta} = (\beta_1,...,\beta_n)$ are the weights with which neurons $1\ldots n$ contribute to the decision. We assumed that the decision is unbiased for a stimulus that does not contain any evidence, or equal amounts of evidence for both choices, implying $\langle D\rangle = 0$. Our convention was that if $D < 0$, choice 1 is elicited, if $D > 0$, choice 2 is initiated.

We further assumed that the neuronal responses can be modeled as a multivariate normal distribution whose means are given by the neuron's tuning functions $\langle r_k\rangle = f_k(s)$. We assumed that the neuronal responses can be correlated and denoted the noise covariance matrix with $\mathbf{C}$. C_{kk} is the response variance of neuron k and C_{jk} is the covariance between the responses of neuron j and neuron k.

Based on these assumptions, we derived the choice-conditioned stimulus distribution $P(r_k \mid D < 0)$ for choice 1, and $P(r_k \mid D > 0)$ for choice 2. Given that

$$P(r_k \mid D<0)P(D<0) = P(r_k, D<0) = P(D<0 \mid r_k)P(r_k) \tag{8}$$

and $P(D<0) = P(D>0) = 1/2$ (assuming unbiased decision-making), it follows that

$$P(r_k \mid D<0) = 2P(r_k)P(D<0 \mid r_k) \tag{9}$$

and similarly for $P(r_k \mid D > 0)$. Given that we assumed the r_k to be normally distributed, D is also normally distributed, as is $D \mid r_k$. Applying the formula for the mean and variance of conditional Gaussians, we found

$$P(D \mid r_k) = \phi\left[D : C_{kk}^{-1}\sum_{j=1}^{n}\beta_j C_{kj}\delta r_k, \sum_{j=1}^{n}\sum_{l=1}^{n}\beta_j\beta_l C_{jl} - C_{kk}^{-1}\left(\sum_{j=1}^{n}\beta_j C_{kj}\right)^2\right] \tag{10}$$

In the above, $\phi[x : \langle x\rangle, \mathrm{var}(x)]$ is the probability density function of the normal distribution and $\delta r_k = r_k - f_k(s)$ is the deviation of the response of neuron k from its mean across all trials and choices. Denoting with Φ the cumulative normal distribution function, we obtained

$$\begin{aligned} P(\delta r_k \mid D<0) &= 2\phi(r_k : f_k(s), C_{kk})\Phi\left(0 : \frac{(\mathbf{C\beta})_k}{C_{kk}}\delta r_k; \boldsymbol{\beta}^{\mathrm{T}}\mathbf{C\beta} - \frac{(\mathbf{C\beta})_k^2}{C_{kk}}\right) \\ &= \frac{2}{\sqrt{C_{kk}}}\phi\left(\frac{\delta r_k}{\sqrt{C_{kk}}} : 0,1\right)\Phi\left(-\frac{(\mathbf{C\beta})_k}{\sqrt{C_{kk}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{C\beta} - (\mathbf{C\beta})_k^2}}\frac{\delta r_k}{\sqrt{C_{kk}}} : 0,1\right) \end{aligned} \tag{11}$$

where

$$(\mathbf{C\beta})_k \equiv \sum_{j=1}^{n}\beta_j C_{kj}$$

$$\boldsymbol{\beta}^{\mathrm{T}}\mathbf{C\beta} \equiv \sum_{j=1}^{n}\sum_{l=1}^{n}\beta_j\beta_l C_{jl}$$

are as described in the main text. Equation (11) represents a skew-normal distribution (generally defined as $P(x) = 2\phi(x)\Phi(\alpha x)$, where α is a scalar shape parameter; **Supplementary Fig. 1a**).

Having derived the general choice-triggered response distribution (equation (11)), we computed the choice probability according to[5]

$$CP_k = \int_{-\infty}^{\infty} d\delta r_k\, P(\delta r_k \mid D>0)\int_{-\infty}^{\delta r_k} d\delta r_k'\, P(\delta r_{k'} \mid D<0) \tag{12}$$

In the following we sketch the solution restricting ourselves to the major steps (for details, see **Supplementary Note**). Defining

$$\alpha = \frac{(\mathbf{C\beta})_k}{\sqrt{C_{kk}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{C\beta} - (\mathbf{C\beta})_k^2}}$$

it follows from equations (11) and (12) that

$$\begin{aligned} CP_k &= 4\int_{-\infty}^{\infty} dx\,\phi(x)\Phi(\alpha x)\int_{-\infty}^{x} dy\,\phi(y)[1-\Phi(\alpha y)] \\ &= 4\left[\int_{-\infty}^{\infty} dx\,\phi(x)\Phi(x)\Phi(\alpha x) - \int_{-\infty}^{\infty} dx\,\phi(x)\Phi(\alpha x)\int_{-\infty}^{x} dy\,\phi(y)\Phi(\alpha y)\right] \end{aligned}$$

where zero mean and unit variance have been omitted from ϕ and Φ for brevity. Partially integrating both terms, we obtained

$$CP_k = \frac{3}{2} - 2\alpha\int_{-\infty}^{\infty} dx\,\phi(\alpha x)\Phi(x)^2 \tag{13}$$

We performed the integral on the right, $F(\alpha) = \int_{-\infty}^{\infty} dx\,\phi(\alpha x)\Phi(x)^2$, by differentiation and integrated to find

$$\begin{aligned} \frac{dF(\alpha)}{d\alpha} &= -\alpha\int_{-\infty}^{\infty} dx\,x^2\phi(\alpha x)\Phi(x)^2 \\ &= -\frac{1}{\alpha}F(\alpha) - \frac{2}{\alpha}\int_{-\infty}^{\infty} dx\,x\phi(\alpha x)\phi(x)\Phi(x) \\ &= -\frac{1}{\alpha}F(\alpha) - \frac{1}{\pi}\frac{1}{\alpha(1+\alpha^2)\sqrt{2+\alpha^2}} \end{aligned} \tag{14}$$

The homogenous part of this differential equation implies $F(\alpha) \propto 1/\alpha$, leading to the ansatz $F(\alpha) = g(\alpha)/\alpha$. Substituting this into equation (14) and integrating yielded

$$g(\alpha) = -\frac{1}{\pi}\arctan\left(\frac{\alpha}{\sqrt{\alpha^2+2}}\right) + c$$

where c is an integration constant. Substituting $g(\alpha)$ back into F, and F into eq. (13), and choosing c appropriately, we arrived at

$$CP_k = \frac{1}{2} + \frac{2}{\pi}\arctan\frac{\alpha}{\sqrt{\alpha^2+2}}$$

which, after substituting in α, yields our central result in equation (1).

Choice probability convention. In our convention, choice probabilities are always reported with respect to choice 1. This means that the choice probabilities can range from 0 to 1, with values smaller than 0.5 indicating that the neural noise for such a neuron is anticorrelated with choice 1. Given that a choice probability of x with respect to choice 1 is equivalent to a choice probability of $1 - x$ with respect to choice 2, experimental papers usually convert all choice probabilities to values between 0.5 and 1.

Details for reconstruction of weights. If we write the read-out weight function $\beta(\phi)$ as linear combination of the eigenfunctions, $v^{(i)}(\phi)$, of the covariance function, $c(\phi_1,\phi_2)$,

$$\beta(\phi) = \sum_i v^{(i)}v^{(i)}(\phi) \tag{15}$$

we find

$$CP(\phi) \approx \frac{1}{2} + \frac{\sqrt{2}}{\pi}\frac{\sum_i v^{(i)}\lambda_i v^{(i)}(\phi)}{\sqrt{c(\phi,\phi)\sum_i v^{(i)2}\lambda_i}} \tag{16}$$

Multiplying this equation by $v^{(k)}(\phi)$, rearranging terms, and integrating over ϕ on both sides yields

$$\frac{\pi}{\lambda_k\sqrt{2}}\int d\phi\left[CP(\phi) - \frac{1}{2}\right]\sqrt{c(\phi,\phi)}v^{(k)}(\phi) \approx \frac{v^{(k)}}{\sqrt{\sum_i v^{(i)2}\lambda_i}} \tag{17}$$

Computing the left-hand side of this equation using the measured choice probabilities and response variabilities, C_{jj}, yields

$$\frac{\pi}{\lambda_k\sqrt{2}}\frac{1}{N}\sum_{j=1}^{N}[CP(\phi_j)-0.5]\sqrt{C_{jj}}\nu^{(k)}(\phi_j)\approx\frac{\nu^{(k)}}{\sqrt{\sum_i \nu^{(i)2}\lambda_i}} \quad (18)$$

The right hand side of equation (17) is the desired coefficient $\nu^{(k)}$, normalized by $\boldsymbol{\beta}^T\mathbf{C}\boldsymbol{\beta}=\sum_i \nu^{(i)2}\lambda_i$. However, as the overall scaling of β, and hence of the coefficients $\nu^{(k)}$, does not affect behavior or choice probabilities, we can define $\boldsymbol{\beta}^T\mathbf{C}\boldsymbol{\beta}\equiv 1$s without loss of generality to obtain $\nu^{(k)}$.

The number of choice probability measurements and their accuracy (number of trials) limits the accuracy with which the left hand side of equation (18) can be determined from experiments. At the same time, the quality of the covariance measurements limits the accuracy with which eigenfunctions and eigenvalues can be computed. Together, this implies that only the coefficients belonging to the eigenfunctions with the largest eigenvalues can be estimated reliably from data. For our simulations, we restricted the reconstruction to the seven eigenfunctions with the largest eigenvalues (**Supplementary Fig. 8**). In our case, the rotational symmetry of the problem implies that, within each pool, the weight profile should be symmetric around the decision axis (here π/2 and 3π/2) and antisymmetric around the pool boundary at π, which is why, for simplicity, we only consider the two of those seven eigenfunctions that obeyed the rotational symmetries of the task (**Supplementary Fig. 8c**,**d**). In principle, the read-out profile used by the brain may not conform to these symmetry requirements and may be biased in some way (for example, to cardinal axes, or as a result of experience from prior tasks). Thus, it would be interesting to compute the coefficients $\nu^{(i)}$ for the other eigenfunctions that do not obey these symmetry requirements and (by boot-strapping the data to get confidence intervals) to test whether they are significantly different from zero.

Simulation details. Our 'ground truth' is a population consisting of 2,048 neurons with diverse tuning curve amplitudes and tuning widths, response variability and correlation coefficients. The tuning curve of each neuron is given by a von-Mises function whose width parameter κ is randomly drawn from a distribution covering values between 0.3 and 5, with a peak at 1.5. Tuning curve amplitudes and baselines are drawn from a Poisson distribution with a mean of 10 and their preferred directions evenly cover the 0…2π range (**Supplementary Fig. 11a–c**). Furthermore, we assumed the neurons' response variability to be proportional to their mean response. To account for the underlying heterogeneity in correlation coefficients, we sampled the correlation matrix from a Wishart distribution around a mean based on linear fits (separate in each quadrant) to empirical data for this task[26] (**Fig. 3a**,**b**). The individual correlation coefficients had a standard deviation of 0.025 (similar to empirical data[39] of 0.03; A. Ecker, personal communication). Note that the observed variability is much higher as a result of measurement noise from the limited numbers of trials (**Supplementary Fig. 8b**). We recorded responses from 128 pairs of neurons in 200 simulated trials each and, from these responses, estimated the correlation coefficients and choice probabilities. The estimated correlation coefficients were used to fit a piece-wise linear correlation function for which we computed eigenvalues and eigenfunctions necessary for the read-out weight reconstruction (**Supplementary Fig. 8a**,**b**). To compute the neuronal sensitivity in **Figure 6b**, we assumed 200 signal trials for each of the two stimuli that were to be distinguished. The statistical power of the tests distinguishing between uniform and optimal read-out based on the relative weight on the first two symmetric eigenfunctions (**Fig. 5c**) and deviation from the sensitivity–choice probability proportionality (**Fig. 6c**), respectively, was quantified by the areas under the ROC curve.

Code. Matlab code implementing the analytical computation of choice probabilities from read-out weights and noise correlations, and the reconstruction of the decoding strategy from choice probabilities and noise correlations, is available at http://bethgelab.org/code/haefner2013/. This material has not been peer reviewed.

doi:10.1038/nn.3309