SUPPLEMENTAL INFORMATION

Cracking the neural code for sensory perception by combining statistics, intervention and behavior

Stefano Panzeri[1,2], Christopher D. Harvey[3], Eugenio Piasini[1], Peter E. Latham[4], & Tommaso Fellin[2,5]

[1] Neural Computation Laboratory, Istituto Italiano di Tecnologia, Rovereto, Italy
[2] Neural Coding Laboratory, Istituto Italiano di Tecnologia, Rovereto, Italy
[3] Department of Neurobiology, Harvard Medical School, Boston, MA, USA
[4] Gatsby Computational Neuroscience Unit, University College London, London, UK
[5] Optical Approaches to Brain Function Laboratory, Istituto Italiano di Tecnologia, Genoa, Italy

SUPPLEMENTAL METHODS

Stimulus information, choice information, and their intersection

In the following subsections, we review the measures that are used to quantify selectivity of neurons to stimulus, choice, and their intersection, and we comment on the strengths and weaknesses of the various measures. For simplicity, we assume that there are two possible stimuli and two possible choices; all of the measures except the one based on signal detection theory readily generalize to more. We will use an *n*-dimensional vector belonging to a set *R*, $\mathbf{r} \in R$, to refer to a neural response quantified by a set of *n* response features, so $r = (r_1, ..., r_n)$. First we consider information about stimulus and choice separately, then we consider their intersection.

Stimulus and choice information

We start by describing various measures to quantify the relationship between response features and external variables. Because this treatment applies to both stimuli and behavioral choices, we use $x \in X = \{1,2\}$ to refer to a generic external correlate; x refers either to stimulus, s (belonging to a set $S = \{1,2\}$), or choice, c (belonging to a set $C = \{1,2\}$).

Fraction correct

A simple quantification of sensory discriminability is the fraction correct – the fraction of times the stimulus decoded from a neural response feature on a single trial matches the actual stimulus on that trial (Quian Quiroga and Panzeri, 2009). Similarly, a simple quantification of choice discriminability is the fraction of choices decoded from a neural response that match the actual choice made by the animal. This measure depends on the choice we make for the decoding algorithm, of which there are many. Probably the most common one is a linear classifier (the decoding and decision boundaries drawn in Figs 3 and S1 are examples of it), which "draws" a linear boundary delimiting the parts of the response space that lead to decoding a particular value of the stimulus or choice. For two stimuli or responses, say x=1 or 2, the boundary is specified by a direction, **w**, and a threshold, θ, such that

$$x = \begin{cases} 1 \text{ if } \mathbf{w} \cdot \mathbf{r} > \theta \\ 2 \text{ if } \mathbf{w} \cdot \mathbf{r} \le \theta. \end{cases} \tag{S1}$$

Other methods, such as Bayesian decoding(Gelman et al., 2014), build a decoding rule that associates each neural response, **r**, with the value of the stimulus or choice, x. Often the decoded value is the one that maximizes the posterior probability,

$$\begin{aligned} x(\mathbf{r}) &= \underset{x}{\operatorname{argmax}}\, P(x \mid \mathbf{r}) \\ P(x \mid \mathbf{r}) &= \frac{P(\mathbf{r} \mid x) P(x)}{P(\mathbf{r})} \end{aligned} \tag{S2}$$

where $P(\mathbf{r}|x)$ is the probability of response $\mathbf{r}$ given x, $P(x)$ is the prior probability of stimulus or choice, and $P(\mathbf{r})$ is the prior probability of response r.

Decoding performance computed as fraction correct has several advantages over other, more complex, measures such as information-theoretic ones (Quian Quiroga and Panzeri, 2009): it is easy to compute, it has a very intuitive interpretation, and it is does not require a large amount of data to estimate accurately. It also has at least two disadvantages relative to information theoretic measures: it does not capture all ways in which a neural response may carry information (the fraction correct may be at chance level – the level one would predict without observing neural activity – even when the neural activity does convey some information about the stimulus or the choice), and it depends on the specific decoding algorithm used for the analysis.

Area under the Receiver Operating Characteristic curve

A measure based on signal detection theory computes the probability that a random sample from the distribution of one stimulus (or choice) is larger than a random sample from the distribution of the other stimulus (or choice). This measure in general requires a one-dimensional response (but see (Haker et al., 2005; Safaai et al., 2013) for attempts to extend it to two-dimensional responses), which we'll take to be $\mathbf{w}\cdot\mathbf{r}$ (in most applications the weight, $\mathbf{w}$, picks out one of the components of $\mathbf{r}$, but this is not necessary). In neuroscience, this measure is known as the neural sensitivity and choice probability for stimulus and choice selectivity respectively; see (Britten et al., 1996; Shadlen et al., 1996). The signal detection theory measure of discriminability of the external variable, x, based on $\mathbf{w}\cdot\mathbf{r}$, is quantified by the Area Under the Receiver Operating Characteristic curve (AUROC, see ref. (Dayan and Abbot, 2001)), which is defined as

$$AUROC = \sum_{\mathbf{r}} p(\mathbf{w}\cdot\mathbf{r} \mid x=2) \sum_{\mathbf{r}':\mathbf{w}\cdot\mathbf{r}'<\mathbf{w}\cdot\mathbf{r}} p(\mathbf{w}\cdot\mathbf{r}' \mid x=1). \quad \text{(S3)}$$

where, as above, x can take on the values 1 or 2. The AUROC can be understood in terms of a trade-off between the false alarm rate (the probability of choosing x=2 when x=1) and the hit rate (the probability of choosing x=2 when x is in fact equal to 2). In mathematical terms, it corresponds to the integral of the hit rate as a function of the false alarm rate, for all possible decision threshold values. AUROC is 0.5 if the conditional distributions of $\mathbf{w}\cdot\mathbf{r}$ given x=1 and x=2 are identical, and increases up to 1 as the two distributions become more and more separated. This measure is closely related to fraction correct under a linear decoder, and so has similar advantages and disadvantages: its advantages are data robustness and ease of interpretability; its disadvantages are that it does not capture all ways in which a neural response may carry information. In addition, its interpretation as a single-trial measure is not as direct as it is for fraction correct or mutual information. That's because AUROC is the probability that the response of a random trial from one stimulus (or choice) is larger than the response in another random trial from the other stimulus (or choice). Turning the AUROC into a single trial measure thus requires the conceptual introduction of an "anti-neuron". Such a neuron responds as if the non-presented stimulus (or choice) had been presented. For instance, if x=1 the anti-neuron responds as if x=2 (i.e. $p(\mathbf{w}\cdot\tilde{\mathbf{r}}|x=1) = p(\mathbf{w}\cdot\mathbf{r}|x=2)$, where $\tilde{\mathbf{r}}$ is the response of the anti-neuron), and if x=2 the anti-neuron responds as if x=1 ($p(\mathbf{w}\cdot\tilde{\mathbf{r}}|x=2) = p(\mathbf{w}\cdot\mathbf{r}|x=1)$). The AUROC then gives the probability that, in any

given trial, $\mathbf{w} \cdot \tilde{\mathbf{r}} > \mathbf{w} \cdot \mathbf{r}$ if x=1 or $\mathbf{w} \cdot \tilde{\mathbf{r}} < \mathbf{w} \cdot \mathbf{r}$ if x=2 (see (Britten et al., 1996)). This concept of antineuron does not necessarily have an immediate biological plausibility. Nevertheless, the AUROC increases monotonically as decoding gets easier, making it a good and often used measure of dependency.

Generalized linear models

An increasingly popular approach is to fit neural responses with generalized linear models. These are models that parametrize the neural response distribution as a function of a linear combination of behavioral and experimental variables – in our case, a linear combination of stimuli and choices. Once the models are fit to data, selectivity to stimuli and choice can be inferred from the weights linking those variables to the neural response (Park et al., 2014; Pillow et al., 2008). Large and statistically significant weights to a given variable imply a strong dependence on that variable. Statistically null weights to a variable imply that the neural response does not depend on it. The advantage of these models is that they have excellent convergence properties, and there are well-developed model regularization tools that allow fitting models to data even when there are a large number of external variables. The second advantage is an important one, as it means these model can be used to study the effect of large numbers of external variables on neural activity (Friedman et al., 2010). A disadvantage is that they make assumptions about the form of the response distribution; if those assumptions are wrong, the model may give misleading results.

Information theoretic quantities

Probably the most general measure of the relationship between the response and the stimulus or choice is the mutual information. Mutual information quantifies, in units of bits, the average reduction of uncertainty about which stimulus was presented (or which choice was taken) based on a single-trial observation of the neural response. Mutual information captures all possible relationships between a neural response and the stimulus or choice, including non-linear ones (Quian Quiroga and Panzeri, 2009; Shannon, 1948). Mutual information, $I(X;\boldsymbol{R})$, between external variable x belonging to set X and neural response $\mathbf{r}$ belonging to set $\boldsymbol{R}$ is defined as

$$I(X;\boldsymbol{R}) = \sum_{\mathbf{r}} \sum_{x} P(x) P(\mathbf{r} \mid x) \log_2 \frac{P(\mathbf{r} \mid x)}{P(r)} \tag{S4}$$

where $P(x)$, $P(\mathbf{r}|x)$ and $P(\mathbf{r})$ were defined above. The mutual information is zero only when the response is independent of x, as in that case no knowledge about x can be gained by observing the response. Unlike other simpler correlation measures, information captures all dependences between the response and the stimulus or the choice. Its main disadvantage is that it is extremely hard to compute from data (Panzeri et al., 2007).

Intersection information

Statistical intersection information

The above measures focus on the stimulus and choice separately. However, as discussed in the main text, they don't provide a direct measure of whether response features useful for decoding the stimulus are also used by the animal to make decisions. Here, we follow (Zuo et al., 2015), to describe a recently developed measure for it, which we refer to as the Intersection Information, denoted *II*. Conceptually, we can think of it either as the amount of

sensory information that is read out in a single trial from a given neural response feature, or the effect on task performance of the sensory information carried by the feature.

The proposal of (Zuo et al., 2015) to empirically quantify intersection information from data tries to capture the contribution of neural features to task performance based on the idea that intersection information should be high when the accuracy of the sensory information carried by the neural feature co-varies with the correctness of the behavioral choice. That is, high intersection information is found when neural response feature, **r**, carries information about both the stimulus and choice, and, importantly, the choice is likely to agree, trial by trial, with the information that the neural response **r** provides about the stimulus. Therefore, a measure of intersection should be based on the probability that a correct behavioral choice co-occurs on a trial-by-trial basis with a correct representation of the stimulus by the neural response. This measure can be computed, from the probability of the animal's choice *c* and the stimulus $\hat{s} = \hat{s}(\mathbf{r})$ decoded from neural activity **r** conditional to the presentation of stimulus *s*:

$$p(\hat{s}, c|s) = \frac{1}{p(s)} \sum_{\mathbf{r}} p(s, \mathbf{r}, \hat{s}, c) = \frac{1}{p(s)} \sum_{\mathbf{r}} p(\hat{s}|\mathbf{r}) p(s, \mathbf{r}, c). \qquad \text{(S5)}$$

Note that the two distributions $p(\hat{s}\,|\,\mathbf{r})$ and $p(s,\mathbf{r},c)$ have slightly different interpretations. The first, $p(\hat{s}\,|\,\mathbf{r})$, depends on the decoding algorithm and so is up to the experimenter; it contains, therefore, assumptions about sensory coding. This probability could be a deterministic decoder, such as Eq. (S1), with *x*=*s*, or it could be probabilistic – either a close approximation to $p(s\,|\,\mathbf{r})$ as measured from data, or a parametric fit to a model. The second, $p(s,\mathbf{r},c)$, must correspond to the true distribution – the one measured from data. The decomposition on the right hand side of Eq. (S5) holds because by construction $\hat{s}$ is assumed to depend exclusively on the neural response r and not on the stimulus s.

To evaluate the statistical significance of intersection information, we have to compare $p(\hat{s},c\,|\,s)$ to the "chance" distribution $p^n(\hat{s},c\,|\,s)$ – the distribution we would obtain under the null hypothesis that there is no relationship between the accuracy of the neural representation of the stimulus in a trial and the correctness of the choice made by the animal in that same trial. This corresponds to a null hypothesis distribution $p^n(\hat{s},c\,|\,s)$ with the same distribution of decoded stimuli as the data (that is, $p^n(\hat{s}\,|\,s) = p(\hat{s}\,|\,s)$) and with the same behavioral performance for each stimulus as the data ($p^n(c\,|\,s) = p(c\,|\,s)$), but for which the decoded stimulus is independent of choice at fixed stimulus:

$$p^n(\hat{s},c\,|\,s) = p(\hat{s}\,|\,s)\,p(c\,|\,s). \qquad \text{(S6)}$$

The null-hypothesis expression in Eq. (S6) reflects the fact that when no stimulus information carried by **r** is used for the task, the probability of the animal making a correct choice does not depend on whether or not the stimulus was decoded correctly on that trial, but depends only on the conditional probability of each choice given the stimulus.

Importantly, $p(\hat{s},c\,|\,s)$ in Eq. (S5) and its null-hypothesis version in Eq. (S6) are both properly normalized probability functions. Thus, we can use these probabilities to define an

intersection measure.

Our simple definition of intersection information *II* is the probability that the stimulus is decoded correctly and the animal makes the correct choice (where, as mentioned above, the correct association between presented stimulus and choice is experimenter-defined, and learned by the animal). In other words, intersection information, *II*, is the probability that the stimulus is decoded correctly given neural features, **r**, and that the correct choice is made on the same trial. Thus, this quantity measures the impact of the neural features on task performance, and it has the following expression:

$$II = \sum_{i=1,2} p(s=i)\,p(\hat{s}=i, c=i \mid s=i) \qquad \text{(S7)}$$

where we assumed, without loss of generality, that the stimuli and choices are numbered so that the correct choice associated with stimulus $s=i$ is choice $c=i$.

The "chance" level for *II* is obtained by substituting p^n (Eq. S6) instead of p in Eq. S7:

$$II^{\mathrm{n}} = \sum_{i=1,2} p(s=i)\,p^n(\hat{s}=i, c=i \mid s=i) = \sum_{i=1,2} p(s=i)\,p(\hat{s}=i \mid s=i)\,p(c=i \mid s=i) \qquad \text{(S8)}$$

A value of *II* higher than chance means that there are more instances of trials with both correct decoding and correct choice than could be expected by chance (thus, chance intersection is the amount of intersection achieved when correctness of choice in a trial does not depend on the correctness of sensory information carried by the features in that trial). Furthermore, *II* is bounded from above by the behavioral and decoding performance, measured respectively as fraction of correct-behavior trials and trials where the stimulus was correctly decoded from the neural feature **r**:

$$\begin{aligned} II &= \sum_{i=1,2} p(s=i)\,p(\hat{s}=i, c=i \mid s=i) \\ &\le \sum_{\hat{s}=1,2}\sum_{i=1,2} p(s=i)\,p(\hat{s}, c=i \mid s=i) = \sum_{i=1,2} p(s=i)\,p(c=i \mid s=i) \end{aligned} \qquad \text{(S9)}$$

and

$$II \le \sum_{c=1,2}\sum_{i=1,2} p(s=i)\,p(\hat{s}=i, c \mid s=i) = \sum_{i=1,2} p(s=i)\,p(\hat{s}=i \mid s=i) \qquad \text{(S10)}$$

Thus, this intersection information is a reasonable quantification of the total impact on task performance of the neural response. A high value requires both high values of sensory information and near-optimal readout (the maximal value of *II* is reached when the sensory code is faultless *and* the readout uses all the sensory information). The values of *II* and its chance level for the three examples presented in Fig. 3 are shown in Fig. S1.

Ref. (Zuo et al., 2015) elaborated that a neural code that affects behavior is also expected to lead the animal to make a behavioral erroneous choice when the stimulus decoded by neural activity is the wrong one (In Ref. (Zuo et al., 2015) these trials were termed the trials carrying misleading sensory information). Thus one possible way to further extend the definition of *II* is to consider separately as an additional quantification (Zuo et al., 2015), not only the sum over trials with correct decoding and correct behavioral choice (i.e., trials with $c=s=\hat{s}=i$ as in Eq (S7)) but also the sum over trials with incorrect decoding and incorrect behavioral choice

(i.e. trials with $s=1, \hat{s}=c=2$, and trials with $s=2, \hat{s}=c=1$). The intersection measure computed over the unfaithful trials is useful to further test the statistical association between sensory information in a neural feature and behavior. In cases when two neural features carry equal amounts of intersection information only on the correctly decoded and behaviorally correct trials, neural features with higher intersection information in incorrectly decoded and behaviorally incorrect trials make a stronger case for a candidate neural code, as these feature show a tighter association with behavioral choice over all trials.

Another normalization for intersection information measures, which was also introduced in (Zuo et al., 2015), is a quantity that we here denote as the fraction of intersection information, shortened as *fII*. It is the fraction of correctly-decoded trials on which the decoded stimulus coincides with that reported by the animal. Unlike *II*, *fII* does not depend on the fraction of times the stimulus is decoded correctly from neural feature $\boldsymbol{r}$; it is given by

$$fII = \sum_{i=1,2} p(s=i)\, p(c=i \mid s=i, \hat{s}=i) \tag{S11}$$

The "chance" level of *fII* is obtained by replacing p in Eq. (S11) with p^n of Eq. (S6) and, as demonstrated by the following equation, simply equals the average fraction of behaviorally correct trials:

$$\begin{aligned} fII^{\mathrm{n}} &= \sum_{i=1,2} p(s=i)\, p^{\mathrm{n}}(c=i \mid s=i, \hat{s}=i) \\ &= \sum_{i=1,2} p(s=i) \frac{p^{\mathrm{n}}(s=i, \hat{s}=i, c=i)}{p^{\mathrm{n}}(s=i, \hat{s}=i)} \\ &= \sum_{i=1,2} p(s=i) \frac{p(s=i)\, p(c=i \mid s=i)\, p(\hat{s}=i \mid s=i)}{p(s=i)\, p(\hat{s}=i \mid s=i)} \\ &= \sum_{i=1,2} p(s=i)\, p(c=i \mid s=i) \end{aligned} \tag{S12}$$

Two cases with the same alignment between decoding and decision boundary but different amounts of stimulus information would therefore have the same value of *fII*, but a different value of *II* (the case with larger stimulus information would give larger *II*). Thus, *fII* is more sensitive to the optimality of the readout – in the linear case, the alignment between the decoding and readout boundaries – than to the total impact of the neural feature **r** on task performance.

These intersection information measures can be used to rank features according to their potential importance for task performance. Importantly, the intersection information is low if a neural response feature has only sensory information and not choice information, or vice versa, or if the sensory information and choice information do not overlap.

Understanding the relationship between the neuroscience question and the measure of intersection is an open area of research. Here we introduced the concept of intersection information from an empirical point of view, and we discussed its practical and conceptual importance for guiding future studies of the neural code. We expect the computational neuroscience community to evaluate this concept with rigor and in detail, and come up with optimal measures of it in the near future.

Interventional intersection information

The intersection quantities defined in a statistical way in the previous sections were designed to be computed from naturally evoked responses. The generalization of these statistical quantities trivially extends to responses generated interventionally. Here we spell this out for the convenience of our readers.

Let **r** be the neural features generated by intervention in one trial, and let *c* be the choice taken by the animal in response to this intervention. In brief, the interventional intersection quantities are obtained from the Eqs. (S7-S12) of the statistical intersection measures by replacing the statistical probability, $p(c|\mathbf{r})$, of choice given neural feature obtained with natural responses with the analogous interventional probability of choice given neural feature **r** obtained under intervention. In the following, we discuss the meaning and implications of different ways of computing intersection information with intervention.

The simplest interventional intersection measure that could be computed from intervention experiments is the interventional fraction of intersection information, *fII*, which (exactly as in the statistical case, Eq. (S11)) is defined simply as the fraction of intervention trials in which the behavioral choice reports the stimulus that would be decoded from the response, **r**, elicited by intervention. However, the interventional *fII*, like its analogous statistical measure, does not take into account whether the stimulus information (that is, fraction of correctly decoded trials) of the considered neural feature is small or large under naturally-evoked conditions. This is a problem if we want to be able to rank, after an interventional experiment, neural features in terms of their contribution to task performance (there could be two neural features that are similarly optimally read out according to *fII*, but one of the features may have higher sensory information and so have a larger impact on behavioral task performance).

To measure an interventional analogue of *II*, we need to consider how likely it is that the evoked pattern, **r**, in natural conditions would appear for each stimulus. Thus, when calculating an interventional *II*, we need to use the distribution $p(\mathbf{r}|s)$ of neural features given the stimulus, *s*, measured under natural conditions. This can be achieved by summing over all tested elicited patterns **r**, and weighting the probabilities of $\hat{s}$ and *c* observed with each value of the interventionally evoked neural feature **r** with their natural probability $p(\mathbf{r}|s)$, as in Eqs. (S7-S12).

This consideration emphasizes that computing the intervention intersection and evaluating the causal impact of a neural code demands a statistical analysis of the probability of naturally occurring patterns during the presentation of stimuli during the task. This is a key point of the framework we propose.

Limitations of measuring separately sensory and readout information without measuring their single trial intersection

To complement the material provided in the main text and in the above Supplemental Information sections, in this section we spell out more examples of the potential dangers of measuring separately sensory and readout information, without measuring their single trial

intersection. In particular, we consider examples of null (chance-level) intersection between sensory and information readout even when the neural features correlate with both choice and stimulus.

One case in which a feature (or set of features) **r** may spuriously appear as both choice-informative and stimulus-informative without truly contributing to the animal's choice and performance is when the choice selectivity of **r** is inherited because **r**, although by itself it does not affect choice, inherits choice selectivity by being correlated with a variable that affects choice (Ince et al., 2012). One possibility is the case plotted in Fig. 3B and S1B. In this case, variable r_2 does not affect choice (the decision boundary is vertical); however r_2 correlates (because of signal correlations) with r_1, which instead affects choice. As a result, in this example r_2 has spurious choice information (as shown in Fig S1B by the fact that the marginal probabilities of r_2 are choice dependent). As detailed in the main text, this spurious choice selectivity can be revealed statistically and interventionally by studying the joint intersection information of the two variables and comparing it to the intersection information carried by each variable alone. Another case when this confound may arise is if the selectivity of **r** to choice appears because **r** depends on the stimulus even if it has no effect on choice, but the choice correlates with the stimulus. This may happen for example if the animal performs the task above chance level (implying that there is a correlation between the presented stimulus and the animal's choice) without relying on the information in the considered features **r**. This confound of spurious choice selectivity cannot be ruled out by measuring separately the neural feature's information about choice and stimulus, see (Ince et al., 2012). However, our measure of intersection information *II* (Eqs. S7,S11) could rule out this confound because the chance level intersection information (Eq S8) corresponds precisely to a "null hypothesis" case of correctness of choice non depending on correctness of feature's decoding (see Eq.S6 for the null hypothesis probability $p^n(\hat{s},c\,|\,s)$). Thus, within the intersection information framework this confounder may be ruled out simply by comparing *II* to its chance level. For traditional sensory and choice information measures, this confounder may be ruled out by conditioning the measure on the stimulus, as this removes the effect of any shared variability between neural features and choice that may be due only to separate covariation of choice and neural features with stimulus (Ince et al., 2012).

A popular method to measure whether sensory information is transmitted to the readout consists in measuring the correlation of the "psychometric" behavioral performance of the animal, for example the fraction of correct discriminations as a function of a stimulus parameter, with the "neurometric" stimulus discriminability obtained by decoding single-trial responses (Newsome et al., 1989; Romo and Salinas, 2003). This measure is extremely useful and it has led to important results, for example about the role of timing in neural coding (Engineer et al., 2008; Luna et al., 2005; Newsome et al., 1989; Romo and Salinas, 2003). However, given that the neurometric to psychometric performance correlation does not consider the within-the-same trial relationship between the sensory signal carried by the neural features and the animals' choice (but rather compares them only across a whole set of trials), it potentially suffers from similar confounders (discussed in the main text) that affect separate measures of choice and stimulus. In Fig. S1A, we show a case with no intersection information where the stimulus discriminability based on the two neural features (r_1,r_2) (i.e. the neurometric performance of features (r_1,r_2)) closely correlates with the psychometric performance of the animal. In this example, features (r_1,r_2) also have significant choice probability (Britten et al., 1996) in the sense that the choice co-varies with the neural features

on a trial-by-trial basis at fixed (or uninformative) stimulus. This situation arises because the behavioral performance is determined by a third neural feature r_3 that has similar stimulus tuning to r_1 and r_2, and fluctuations along the dimension of (r_1, r_2) that influences behavior are statistically independent from those along the dimension which encodes the stimulus. In this example, however, features r_1, r_2 have null (chance-level) intersection information because there are no noise correlations between all the features (so r_3 is independent of r_1 and r_2 conditioned on the stimulus). This implies that the single trial fluctuations of the stimulus information in r_1, r_2 do not influence choice in the same trial.

To illustrate this quantitatively, we build on the tasks we described in the main text. There are two stimuli that lead to different response distributions. However, we add another parameter, called stimulus signal intensity (shortened to signal intensity) and denoted ρ in the following equations, that controls task difficulty by spreading out or compressing the response distributions (Fig. S1A$_3$-A$_5$). In the green vs blue stimulus exemplified in our paper, signal intensity could be the contrast of the blue or green stimulus with respect to background, so that zero signal intensity mean that the stimulus is invisible from the background and 100% signal intensity means that the stimulus is very well visible from the background. We'll use the convention that the identity of the stimulus is encoded in the sign of ρ: say $\rho<0$ for the green stimulus, $s=1$, and $\rho>0$ for the blue stimulus, $s=2$, or

$$s = \vartheta(\rho) + 1$$

where $\vartheta(\cdot)$ is the Heaviside step-function. We consider three neural response features r_1, r_2 and r_3 (Fig. S1A$_1$), that may represent, for instance, the time of first spike of two neurons (r_1 and r_2) and their total firing rate (r_3). We assume, for concreteness, that the neural response to a stimulus s (with intensity ρ) is given by the Gaussian conditional probability distribution

$$p(\mathbf{r} \mid \rho) = p(r_1, r_2, r_3 \mid \rho) = \mathsf{N}\ (\boldsymbol{\mu}(\rho), \boldsymbol{\Sigma}) = \mid 2\pi\boldsymbol{\Sigma} \mid^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}(\rho)) \cdot \boldsymbol{\Sigma}^{-1} \cdot (\mathbf{r}-\boldsymbol{\mu}(\rho))\right\}$$

where

$$\boldsymbol{\mu}(s) = (\rho, \rho, \rho) \tag{S13}$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \frac{\sigma_+^2 + \sigma_-^2}{4} & \frac{\sigma_+^2 - \sigma_-^2}{4} & 0 \\ \frac{\sigma_+^2 - \sigma_-^2}{4} & \frac{\sigma_+^2 + \sigma_-^2}{4} & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix} \tag{S14}$$

are the mean and the covariance matrix of the distribution, respectively, and σ_+, σ_- and σ_3 are arbitrary parameters controlling sensory encoding noise. This immediately implies that

$$p(r_1 \mid \rho) = p(r_2 \mid \rho) = \mathsf{N}\left(\rho, \frac{\sigma_+^2 + \sigma_-^2}{4}\right)$$

$$p(r_3 \mid \rho) = \mathsf{N}\left(\rho, \sigma_3^2\right)$$

So r_1, r_2 and r_3 all have similar stimulus tuning, and r_1 and r_2 share noise correlations. If we define

$$r_+ = r_1 + r_2$$
$$r_- = r_1 - r_2$$

we have that r_+ and r_- are conditionally independent given the stimulus, i.e. $p(r_+, r_- \mid \rho) = p(r_+ \mid \rho)p(r_- \mid \rho)$, and

$$\begin{aligned} p(r_+ \mid \rho) &= \mathsf{N}\left(2\rho, \sigma_+^2\right) \\ p(r_- \mid \rho) &= \mathsf{N}\left(0, \sigma_-^2\right) \end{aligned} \tag{S15}$$

Now, we suppose that the binary behavioral choice of the animal is given by

$$c(\mathbf{r}) = \vartheta(r_- + r_3) + 1$$

where *c=1* represents "left choice" and *c=2* represents "right choice". The two neural features (r_1, r_2) have higher-than-chance choice information and choice probability, as fluctuations in r_- will bias the choice on a trial-by-trial basis at fixed stimulus or for an uninformative stimulus with $\rho = 0$.

From the definitions above, and considering that $(r_- + r_3)$ N $(\rho, \sigma_-^2 + \sigma_3^2)$, we can compute the probability of a possible choice (c=2) given the stimulus:

$$\begin{aligned} p(c=2 \mid \rho) &= 1 - p(r_- + r_3 < 0 \mid \rho) = 1 - \int_{-\infty}^{0} \frac{1}{\sqrt{2(\sigma_-^2 + \sigma_3^2)\pi}} e^{-\frac{(x-\rho)^2}{2(\sigma_-^2+\sigma_3^2)}} dx \\ &= \Phi\left(\frac{\rho}{\sqrt{\sigma_-^2 + \sigma_3^2}}\right) \end{aligned} \tag{S16}$$

where $\Phi(\cdot)$ is the cumulative Gaussian function. Assuming that $c = 2$ is the correct choice for $\rho > 0$, Eq. (S16) gives the probability that the animal performs correctly, i.e. the psychometric performance of the animal in the task (Fig. S1A$_2$).

Using the same approach, we can compute the neurometric performance of the (r_1, r_2) neural features, defined as the probability of correct stimulus decoding using an ideal decoder. If we assume the green and blue stimuli to be equiprobable ($p(\rho < 0) = p(\rho > 0) = 1/2$), then by a symmetry argument the optimal decoder is that which operates along the sensory boundary $r_+ = 0$ indicated in Figure S1A$_{3\text{-}5}$: $\hat{s} = \vartheta(r_+) + 1$. The probability of correct decoding can be then computed directly from Eq. (S15):

$$p(r_+ < 0 \mid \rho < 0) = p(r_+ > 0 \mid \rho > 0) = \Phi\left(\frac{2\rho}{\sigma_+}\right) \tag{S17}$$

By comparing Eqs. (S16) and (S17), it is apparent that if

$$\sigma_+ > 2\sigma_- \qquad , \qquad \sigma_3 = \sqrt{\frac{\sigma_+^2}{4} - \sigma_-^2}$$

then the neurometric curve for the (r_1, r_2) code coincides with the psychometric curve of the experiment (Fig. S1A$_2$), even though the intersection information of the neural features (r_1, r_2) is at chance level, as the faithfulness of the neural representation of the stimulus is

conditionally independent of the choice given the stimulus (see Eqs. (S7) and (S8)). Indeed, the faithfulness of stimulus encoding only depends on r_+, while the behavioral choice only depends on $r_- + r_3$, and $p(r_+, r_- + r_3 \mid \rho) = p(r_+ \mid \rho)p(r_- + r_3 \mid \rho)$.

Patterned illumination to causally test hypothesis on the intersection between sensory information and readout

To be informative about the neural code, ideally interventional approaches should achieve cellular resolution and high temporal precision in large subpopulations of cells several hundred microns into the brains of mammals (if complex behaviors are to be investigated, rodents or non-human primates models must be used). This is especially important for directly testing hypotheses about the relevance of a particular neural feature (e.g., spike timing or spike count) in particular subsets of neurons. In experimental animal models, optogenetics (Boyden et al., 2005; Lima and Miesenbock, 2005; Nagel et al., 2003; Zemelman et al., 2002; Zhang et al., 2007; Zhang et al., 2010) has become the technique of choice to perturb electrical activity in genetically-targeted cellular subpopulations. Most functional optogenetic studies in living animals have so far used the wide field approach as in Fig. 8C (Adamantidis et al., 2007; Beltramo et al., 2013; Gradinaru et al., 2009; Kravitz et al., 2010; Tsai et al., 2009; Wimmer et al., 2015), which does not allow high spatial resolution within the illuminated region. However, recent optical developments now allow precise spatial targeting (Andrasfalvy et al., 2010; Baker et al., 2016; Papagiakoumou et al., 2010), an approach that is called patterned illumination; see Fig. 8D (Bovetti and Fellin, 2015). Patterned illumination is an umbrella term, and includes different approaches (see below) to deliver light to precise spatial locations. When combined with the light-sensitive optogenetic actuators, patterned illumination can reach near cellular resolution in perturbing electrical activity (Packer et al., 2015; Papagiakoumou et al., 2010; Rickgauer et al., 2014), thus promising to be a powerful tool for investigating the neural code driving behavior. Importantly, patterned illumination has recently been combined with laser scanning functional imaging *in vivo*, providing a unique all-optical tool for reading and perturbing neuronal circuits (Carrillo-Reid et al., 2016; Packer et al., 2015; Rickgauer et al., 2014; Szabo et al., 2014). We will briefly describe here the main technical advancements that have been developed to achieve patterned illumination in the mammalian brain, and discuss their main advantages and limitations. A more technical description of the techniques underlying patterned illumination and their combination with light-sensitive opsin actuators can be found in (Bovetti and Fellin, 2015; Emiliani et al., 2015; Grosenick et al., 2015).

In general, patterned illumination can be performed in combination with both single- (Lutz et al., 2008; Szabo et al., 2014) and two-photon excitation (Andrasfalvy et al., 2010; Packer et al., 2012; Papagiakoumou et al., 2010; Papagiakoumou et al., 2013). Although single-photon patterned illumination might present some advantages for stimulation with fast refresh rates (> 1 kHz using, for example, digital micromirror devices), and is compatible with the excitation of most available opsins, it is unlikely to achieve single-cell resolution in deep regions of the mammalian brain. That's because out-of-focus light activates cellular structures (cell bodies or processes) above and below the target neuron. In addition, scattering limits the applicability of single-photon patterned illumination in turbid mammalian brain tissues. In contrast, two-photon patterned illumination effectively restricts opsin activation in the axial direction (Helmchen and Denk, 2005), assuring cellular resolution hundreds of microns deep within the brain tissue. Patterned two-photon optogenetic illumination can be performed by scanning a diffraction limited spot over a given region of interest (Carrillo-Reid

et al., 2016; Mohanty et al., 2008; Prakash et al., 2012; Rickgauer and Tank, 2009), by providing simultaneous illumination on extended shapes in combination with temporal focusing (Andrasfalvy et al., 2010; Papagiakoumou et al., 2010), or by a combination of light patterning and scanning (Packer et al., 2012; Packer et al., 2015; Rickgauer et al., 2014).

In the scanning approach, cells located within a large field of view (e.g., ~300 μm x 300 μm with a 40X objective or ~600 μm x 600 μm with a 20X objective), potentially containing hundreds of neurons, can be individually addressed. However, this approach does not allow the simultaneous illumination of different cells, and is limited in its time resolution because the sequential scanning mode takes time to address all the target cells. The use of acousto-optic deflectors (Huang et al., 2016; Nadella et al., 2016) may decrease the time necessary to move from one location to the other, but efficient manipulation of the target neuron depends, among other things, on the photo-current rise-time, and therefore on the illumination dwell time. If long dwell times are needed to obtain efficient opsin activation, fast scanning methods might not represent the ultimate solution for stimulating many cells in short time windows. Regardless of these limitations, it has been suggested that using optimized spiral scanning approaches with small dimension galvanometric mirrors and activation of the excitatory opsin C1V1 (Yizhar et al., 2011), approximately 50 neurons can be sequentially addressed in 100 ms (Grosenick et al., 2015).

Patterned illumination using extended two-photon shapes (for example, using liquid crystal spatial light modulators, LC-SLMs) (Dal Maschio et al., 2010; Nikolenko et al., 2008; Papagiakoumou et al., 2010) leads to the simultaneous illumination of larger sample areas. Compared to the scanning of diffraction limited spots, it might potentially be more effective in driving neural cells suprathreshold because it allows the simultaneous illumination of a larger portion of the target cell and thus it leads to the synchronous activation of a higher number of light-sensitive molecules. Moreover, in most configurations this method allows truly simultaneous illumination of multiple neurons. The main limitations include: the addressable area within the field of view is smaller than that of the scanning approach, the number of cells that can be simultaneously illuminated is limited by the total available laser power and tissue heating (Podgorski and Ranganathan, 2016). When series of different patterns need to be projected, the refresh rate of current LC-SLMs is limited (in the order of 60-500 Hz). Although a direct demonstration of the applicability of this technology (using two-photon excitation) to stimulate cells in living mammals still awaits experimental validation, based on published work in brain slices (Begue et al., 2013; Papagiakoumou et al., 2010) it is reasonable to hypothesize that about 10 neurons could be simultaneously stimulated in less than 40 ms when channlerhodopsin-2 is used. The use of other opsins (e.g., ReaChR) combined with low repetition rate laser sources may increase the number of addressable cells while minimizing the latency to action potential (AP) discharge and the AP jitter (Chaigneau et al., 2016).

The combined scanning mirrors and LC-SLM patterned illumination approach might achieve activation of multiple neurons in large fields of views. For example the LC-SLM could be used to shape two-photon light into an extended area corresponding to the dimension of a cell body of a neuron, and galvanometric mirrors could be used to deflect this shape over multiple cells. In a similar way, an extended disk of two-photon excitation could be moved across different neurons (Rickgauer et al., 2014). Alternatively, an LC-SLM can be used to project small excitation spots centered on multiple cells and the galvanometric mirrors could be used to scan the spots on the extended area corresponding to the cell body (Packer et al., 2012;

Packer et al., 2015). Using this approach, 10-20 neurons have been simultaneously stimulated in 11-34 ms (Packer et al., 2015).

To summarize, multiple approaches have been proposed to perform patterned two-photon illumination with near cellular resolution in living rodents. Current experimental approaches can manipulate *in vivo* a relatively small number (few tens) of cells with temporal resolution of few ms (Bovetti and Fellin, 2015; Emiliani et al., 2015; Grosenick et al., 2015). Much effort is currently devoted to combining patterned illumination with neurophysiological measurements in behavioral experiments, but the validity of this approach still awaits experimental demonstration. For example, it is still an open question whether stimulating a limited number of neurons will be sufficient to drive a behavioral response. Success in this task will most likely go through the optimization of stimulation protocols and the development of new technical solutions for efficiently manipulating the activity of hundreds to thousands of cells in three dimensions while maintaining high spatial and temporal resolution (see also main text).

Details of the simulations implemented in this article

The simulations in Figs 3 and 5 and in Fig. S1B were implemented by generating, for each of the two simulated stimuli s=1 and s=2, points in the r_1, r_2 space according to a Gaussian distribution $\mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with covariance matrix

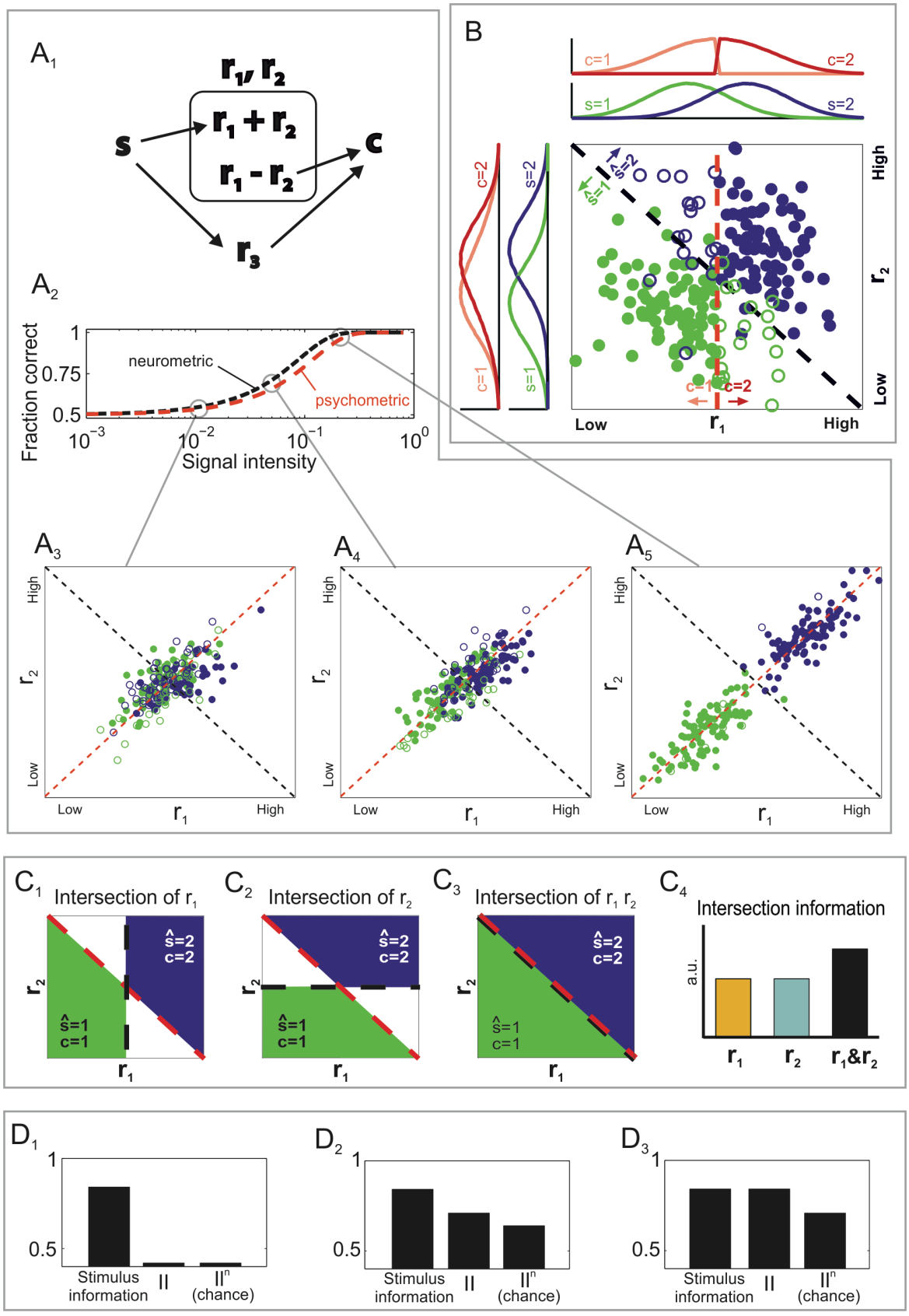
$$\boldsymbol{\Sigma} = \begin{pmatrix} 0.2 & -0.005 \\ -0.005 & 0.2 \end{pmatrix}$$

and mean vector $\boldsymbol{\mu}$ = (0.4, 0.4) for s=1 and $\boldsymbol{\mu}$ = (0.6, 0.6) for s=2. The boxes in the two-dimensional plots of the (R_1, R_2) space in Figs. 3,4 and 6 have axes that span the range between 0 and 1 for each of the two neural features r_1,r_2. The simulations plotted n=100 trials per stimulus in the (R_1, R_2) plane, but the marginal probabilities along the r_1 *and* r_2 axes of Fig. S1B were computed with n=10^6 simulated trials per stimulus.

The simulations in Fig.S1A were generated according to the distribution for (r_1, r_2) defined by Equations S13 and S14, with $\sigma_+ = 0.18$, $\sigma_- = 0.07$ and $\sigma_3 = 0.1$, and s set to either ±0.01 ($S1A_3$), ±0.06 ($S1A_4$) and ±0.2 ($S1A_5$). These parameters were chosen so that the neurometric and psychometric functions plotted in Fig. $S1A_2$ did not completely overlap, for display purposes. The boxes in the two-dimensional plots of the (r_1, r_2) space in Figure S1A have axes that span the range between -0.5 and 0.5 for each of the two neural features r_1 and r_2. As in Figure 3 and Figure S1B, 100 trials per stimulus were plotted.

Matlab code for the generation of these figures is available through Zenodo and GitHub (https://doi.org/10.5281/zenodo.191810).

Supplemental Figure S1



Supplemental Figure S1 (related to Figure 3): further illustrations of intersection information.

A) example of a two features (r_1, r_2) with null (chance-level) intersection information that has significant choice probability and whose neurometric function correlates well with the experiment's psychometric function. The signal intensity of the stimulus is varied parametrically. A_1) Simple schematic illustrating the dependence between stimulus, neural response and behavioral choice. The stimulus is encoded in both r_1 and r_2 as well as in a third feature r_3, and the behavioral readout is based on a combination of $r_1 - r_2$ and r_3 (see text for details). The stimulus tuning (trial-averaged response) of r_1 and r_2 is identical, and similar to that of r_3. A_2) Comparison of the neurometric curve of the two features (r_1, r_2) (black; defined as the probability of correct decoding by an ideal stimulus decoder as a function of signal intensity, using r_1 and r_2; Eq. (S17) and the psychometric curve for the experiment (red; defined as the probability of correct behavioral choice as a function of signal intensity; Eq. S16). $A_{3\text{-}5}$) scatter plot of neural responses for different values of signal intensity, generated according to the distributions defined by Eqs. (S13) and (S14). Graphical conventions are as in Fig. $3A_1$, $3B_1$ and $3C_1$. Dashed black and red lines represent the sensory and decision boundaries, respectively. The region below the sensory boundary corresponds to responses that are decoded correctly from features (r_1, r_2) if the green stimulus is shown; the region above the sensory boundary corresponds to responses that are decoded correctly if the blue stimulus is shown. Filled circles correspond to correct behavioral choices; open circles to wrong choices. As the stimulus signal intensity increases (from ±0.01, to ±0.06 and to ±0.02 respectively in Figure $S1A_3$, $S1A_4$ and $S1A_5$), responses to green and blue stimuli become further apart, and the number of error trials decreases. Notice, however, that there is no single-trial link between the neural representation of a stimulus and the behavioral choice, as the probability of behavioral error in a given trial is always unrelated to its position relative to the sensory boundary. **B)** This figure, which is identical to Fig. $3B_1$ with the addition of marginal probabilities of individual features, is a scatterplot of simulated neural responses to two stimuli, *s=1,2* (corresponding to green and blue dots, respectively). The lines along the axes of the 2-D scatterplot represent the 1-D marginal projections of stimulus- and choice-fixed probabilities of r_2 and r_1, respectively. In this example (which is analogous to that in figure $3B_1$), the decision depends only on r_1, but r_2 also possesses choice selectivity by virtue of its correlation with r_1, as can be seen from the marginal plots on the left. **C)** Sketch showing why, in the case of Fig. 3C, when feature r_1 and r_2 carry complementary stimulus information and the decoder is sensitive to both r_1 and r_2, the intersection information carried by the joint combination of features is larger than the intersection information carried by either alone. The right panel illustrates this by showing histograms of intersection information for individual features (left two) and their joint combination (right). The three panels on the left plot with solid colors (coded with the stimulus color) the regions of the (r_1, r_2) space that contribute to the intersection information if the sensory stimulus is decoded with r_1 only (first from left), with r_2 only (center), and jointly with (r_1, r_2) (right). The larger the colored areas, the larger the intersection information. Decoding with both features maximizes the areas with congruent stimulus and choice information (no "white" areas that do not contribute to intersection). Decoding with either feature alone leads to areas of the (r_1, r_2) plane that cannot contribute to intersection because in these areas there is a mismatch between the decoded stimulus and the choice. The mismatch is indicated by regions (white areas in the feature plane) where the single-feature decoder wrongly decodes the stimulus, due to its failure to consider all the complementary stimulus information in the joint features. **D)** Intersection information values computed for the examples in Figure 3 using Eq. (S7),

compared with “chance” intersection information (Eq. (S8)) and “stimulus information”, quantified as the fraction of trials correctly decoded by the ideal linear stimulus decoder (sensory boundary).

References

Adamantidis, A.R., Zhang, F., Aravanis, A.M., Deisseroth, K., and de Lecea, L. (2007). Neural substrates of awakening probed with optogenetic control of hypocretin neurons. Nature 450, 420-424.

Andrasfalvy, B.K., Zemelman, B.V., Tang, J.Y., and Vaziri, A. (2010). Two-photon single-cell optogenetic control of neuronal activity by sculpted light. Proceedings of the National Academy of Sciences of the United States of America 107, 11981-11986.

Baker, C.A., Elyada, Y.M., Parra, A., and Bolton, M.M. (2016). Cellular resolution circuit mapping with temporal-focused excitation of soma-targeted channelrhodopsin. eLife 5.

Begue, A., Papagiakoumou, E., Leshem, B., Conti, R., Enke, L., Oron, D., and Emiliani, V. (2013). Two-photon excitation in scattering media by spatiotemporally shaped beams and their application in optogenetic stimulation. Biomedical optics express 4, 2869-2879.

Beltramo, R., D'Urso, G., Dal Maschio, M., Farisello, P., Bovetti, S., Clovis, Y., Lassi, G., Tucci, V., De Pietri Tonelli, D., and Fellin, T. (2013). Layer-specific excitatory circuits differentially control recurrent network dynamics in the neocortex. Nature neuroscience 16, 227-234.

Bovetti, S., and Fellin, T. (2015). Optical dissection of brain circuits with patterned illumination through the phase modulation of light. J Neurosci Meth 241, 66-77.

Boyden, E.S., Zhang, F., Bamberg, E., Nagel, G., and Deisseroth, K. (2005). Millisecond-timescale, genetically targeted optical control of neural activity. Nature neuroscience 8, 1263-1268.

Britten, K.H., Newsome, W.T., Shadlen, M.N., Celebrini, S., and Movshon, J.A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis Neurosci 13, 87-100.

Carrillo-Reid, L., Yang, W., Bando, Y., Peterka, D.S., and Yuste, R. (2016). Imprinting and recalling cortical ensembles. Science 353, 691-694.

Chaigneau, E., Ronzitti, E., Gajowa, M.A., Soler-Llavina, G.J., Tanese, D., Brureau, A.Y., Papagiakoumou, E., Zeng, H., and Emiliani, V. (2016). Two-Photon Holographic Stimulation of ReaChR. Front Cell Neurosci 10, 234.

Dal Maschio, M., Difato, F., Beltramo, R., Blau, A., Benfenati, F., and Fellin, T. (2010). Simultaneous two-photon imaging and photo-stimulation with structured light illumination. Optics express 18, 18720-18731.

Dayan, P., and Abbot, L.F. (2001). Theoretical Neuroscience (Cambridge, MA: The MIT Press).

Emiliani, V., Cohen, A.E., Deisseroth, K., and Hausser, M. (2015). All-Optical Interrogation of Neural Circuits. The Journal of neuroscience : the official journal of the Society for Neuroscience 35, 13917-13926.

Engineer, C.T., Perez, C.A., Chen, Y.H., Carraway, R.S., Reed, A.C., Shetake, J.A., Jakkamsetti, V., Chang, K.Q., and Kilgard, M.P. (2008). Cortical activity patterns predict speech discrimination ability. Nature neuroscience 11, 603-608.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of statistical software 33, 1-22.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D. (2014). Bayesian Data Analysis, 3rd edn (Boca Raton (FL): Chapman & Hall).

Gradinaru, V., Mogri, M., Thompson, K.R., Henderson, J.M., and Deisseroth, K. (2009). Optical deconstruction of parkinsonian neural circuitry. Science 324, 354-359.

Grosenick, L., Marshel, J.H., and Deisseroth, K. (2015). Closed-loop and activity-guided optogenetic control. Neuron 86, 106-139.

Haker, S., Wells, W.M., 3rd, Warfield, S.K., Talos, I.F., Bhagwat, J.G., Goldberg-Zimring, D., Mian, A., Ohno-Machado, L., and Zou, K.H. (2005). Combining classifiers using their receiver operating characteristics and maximum likelihood estimation. Medical image computing and computer-assisted intervention : MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention 8, 506-514.

Helmchen, F., and Denk, W. (2005). Deep tissue two-photon microscopy. Nat Methods 2, 932-940.

Huang, L., Ung, K., Garcia, I., Quast, K.B., Cordiner, K., Saggau, P., and Arenkiel, B.R. (2016). Task Learning Promotes Plasticity of Interneuron Connectivity Maps in the Olfactory Bulb. The Journal of neuroscience : the official journal of the Society for Neuroscience 36, 8856-8871.

Ince, R.A., Mazzoni, A., Bartels, A., Logothetis, N.K., and Panzeri, S. (2012). A novel test to determine the significance of neural selectivity to single and multiple potentially correlated stimulus features. J Neurosci Methods 210, 49-65.

Kravitz, A.V., Freeze, B.S., Parker, P.R., Kay, K., Thwin, M.T., Deisseroth, K., and Kreitzer, A.C. (2010). Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. Nature 466, 622-626.

Lima, S.Q., and Miesenbock, G. (2005). Remote control of behavior through genetically targeted photostimulation of neurons. Cell 121, 141-152.

Luna, R., Hernandez, A., Brody, C.D., and Romo, R. (2005). Neural codes for perceptual discrimination in primary somatosensory cortex. Nature neuroscience 8, 1210-1219.

Lutz, C., Otis, T.S., DeSars, V., Charpak, S., DiGregorio, D.A., and Emiliani, V. (2008). Holographic photolysis of caged neurotransmitters. Nat Methods 5, 821-827.

Mohanty, S.K., Reinscheid, R.K., Liu, X., Okamura, N., Krasieva, T.B., and Berns, M.W. (2008). In-depth activation of channelrhodopsin 2-sensitized excitable cells with high spatial resolution using two-photon excitation with a near-infrared laser microbeam. Biophysical journal 95, 3916-3926.

Nadella, K.M., Ros, H., Baragli, C., Griffiths, V.A., Konstantinou, G., Koimtzis, T., Evans, G.J., Kirkby, P.A., and Silver, R.A. (2016). Random-access scanning microscopy for 3D imaging in awake behaving animals. Nat Methods 13, 1001-1004.

Nagel, G., Szellas, T., Huhn, W., Kateriya, S., Adeishvili, N., Berthold, P., Ollig, D., Hegemann, P., and Bamberg, E. (2003). Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. Proceedings of the National Academy of Sciences of the United States of America 100, 13940-13945.

Newsome, W.T., Britten, K.H., and Movshon, J.A. (1989). Neuronal correlates of a perceptual decision. Nature 341, 52-54.

Nikolenko, V., Watson, B.O., Araya, R., Woodruff, A., Peterka, D.S., and Yuste, R. (2008). SLM Microscopy: Scanless Two-Photon Imaging and Photostimulation with Spatial Light Modulators. Frontiers in neural circuits 2, 5.

Packer, A.M., Peterka, D.S., Hirtz, J.J., Prakash, R., Deisseroth, K., and Yuste, R. (2012). Two-photon optogenetics of dendritic spines and neural circuits. Nat Methods 9, 1202-1205.

Packer, A.M., Russell, L.E., Dalgleish, H.W.P., and Hausser, M. (2015). Simultaneous all-optical manipulation and recording of neural circuit activity with cellular resolution in vivo. Nat Methods 12, 140-U186.

Panzeri, S., Senatore, R., Montemurro, M.A., and Petersen, R.S. (2007). Correcting for the sampling bias problem in spike train information measures. J Neurophysiol 98, 1064-1072.

Papagiakoumou, E., Anselmi, F., Begue, A., de Sars, V., Gluckstad, J., Isacoff, E.Y., and Emiliani, V. (2010). Scanless two-photon excitation of channelrhodopsin-2. Nat Methods 7, 848-U117.

Papagiakoumou, E., Bègue, A., Leshem, B., Schwartz, O., Stell, B.M., Bradley, J., Oron, D., and Emiliani, V. (2013). Functional patterned multiphoton excitation deep inside scattering tissue. Nature Photonics 4, 274-278.

Park, I.M., Meister, M.L., Huk, A.C., and Pillow, J.W. (2014). Encoding and decoding in parietal cortex during sensorimotor decision-making. Nature neuroscience 17, 1395-1403.

Pillow, J.W., Shlens, J., Paninski, L., Sher, A., Litke, A.M., Chichilnisky, E.J., and Simoncelli, E.P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 454, 995-999.

Podgorski, K., and Ranganathan, G. (2016). Brain heating induced by near-infrared lasers during multiphoton microscopy. J Neurophysiol 116, 1012-1023.

Prakash, R., Yizhar, O., Grewe, B., Ramakrishnan, C., Wang, N., Goshen, I., Packer, A.M., Peterka, D.S., Yuste, R., Schnitzer, M.J., and Deisseroth, K. (2012). Two-photon optogenetic toolbox for fast inhibition, excitation and bistable modulation. Nat Methods 9, 1171-1179.

Quian Quiroga, R., and Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. Nat Rev Neurosci 10, 173-185.

Rickgauer, J.P., Deisseroth, K., and Tank, D.W. (2014). Simultaneous cellular-resolution optical perturbation and imaging of place cell firing fields. Nature neuroscience 17, 1816-1824.

Rickgauer, J.P., and Tank, D.W. (2009). Two-photon excitation of channelrhodopsin-2 at saturation. Proceedings of the National Academy of Sciences of the United States of America 106, 15025-15030.

Romo, R., and Salinas, E. (2003). Flutter discrimination: neural codes, perception, memory and decision making. Nat Rev Neurosci 4, 203-218.

Safaai, H., von Heimendahl, M., Sorando, J.M., Diamond, M.E., and Maravall, M. (2013). Coordinated population activity underlying texture discrimination in rat barrel cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 33, 5843-5855.

Shadlen, M.N., Britten, K.H., Newsome, W.T., and Movshon, J.A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. The Journal of neuroscience : the official journal of the Society for Neuroscience 16, 1486-1510.

Shannon, C.E. (1948). A mathematical theory of communication. AT&T Tech J 27, 379-423.

Szabo, V., Ventalon, C., De Sars, V., Bradley, J., and Emiliani, V. (2014). Spatially Selective Holographic Photoactivation and Functional Fluorescence Imaging in Freely Behaving Mice with a Fiberscope. Neuron 84, 1157-1169.

Tsai, H.C., Zhang, F., Adamantidis, A., Stuber, G.D., Bonci, A., de Lecea, L., and Deisseroth, K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. Science 324, 1080-1084.

Wimmer, R.D., Schmitt, L.I., Davidson, T.J., Nakajima, M., Deisseroth, K., and Halassa, M.M. (2015). Thalamic control of sensory selection in divided attention. Nature 526, 705-709.

Yizhar, O., Fenno, L.E., Prigge, M., Schneider, F., Davidson, T.J., O'Shea, D.J., Sohal, V.S., Goshen, I., Finkelstein, J., Paz, J.T.*, et al.* (2011). Neocortical excitation/inhibition balance in information processing and social dysfunction. Nature 477, 171-178.

Zemelman, B.V., Lee, G.A., Ng, M., and Miesenbock, G. (2002). Selective photostimulation of genetically chARGed neurons. Neuron 33, 15-22.

Zhang, F., Aravanis, A.M., Adamantidis, A., de Lecea, L., and Deisseroth, K. (2007). Circuit-breakers: optical technologies for probing neural signals and systems. Nat Rev Neurosci 8, 577-581.

Zhang, F., Gradinaru, V., Adamantidis, A.R., Durand, R., Airan, R.D., de Lecea, L., and Deisseroth, K. (2010). Optogenetic interrogation of neural circuits: technology for probing mammalian brain structures. Nature protocols 5, 439-456.

Zuo, Y.F., Safaai, H., Notaro, G., Mazzoni, A., Panzeri, S., and Diamond, M.E. (2015). Complementary Contributions of Spike Timing and Spike Rate to Perceptual Decisions in Rat S1 and S2 Cortex. Current Biology 25, 357-363.