Neuron
Perspective

Cracking the Neural Code for Sensory Perception by Combining Statistics, Intervention, and Behavior

Stefano Panzeri,[1,2,6,*] Christopher D. Harvey,[3,*] Eugenio Piasini,[1] Peter E. Latham,[4] and Tommaso Fellin[2,5,*]
[1]Neural Computation Laboratory
[2]Neural Coding Laboratory
Istituto Italiano di Tecnologia, 38068 Rovereto, Italy
[3]Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA
[4]Gatsby Computational Neuroscience Unit, University College London, London, W1T 4JG, UK
[5]Optical Approaches to Brain Function Laboratory, Istituto Italiano di Tecnologia, 16163 Genoa, Italy
[6]Lead Contact
*Correspondence: stefano.panzeri@iit.it (S.P.), christopher_harvey@hms.harvard.edu (C.D.H.), tommaso.fellin@iit.it (T.F.)
http://dx.doi.org/10.1016/j.neuron.2016.12.036

The two basic processes underlying perceptual decisions—how neural responses encode stimuli, and how they inform behavioral choices—have mainly been studied separately. Thus, although many spatiotemporal features of neural population activity, or "neural codes," have been shown to carry sensory information, it is often unknown whether the brain uses these features for perception. To address this issue, we propose a new framework centered on redefining the neural code as the neural features that carry sensory information used by the animal to drive appropriate behavior; that is, the features that have an intersection between sensory and choice information. We show how this framework leads to a new statistical analysis of neural activity recorded during behavior that can identify such neural codes, and we discuss how to combine intersection-based analysis of neural recordings with intervention on neural activity to determine definitively whether specific neural activity features are involved in a task.

Introduction

To survive, organisms must both accurately represent stimuli in the outside world and use that representation to generate beneficial behavioral actions. Historically, these two processes—the mapping from stimuli to neural responses and the mapping from neural activity to behavior—have mainly been treated separately. Of the two, the former has received the most attention. Often referred to as the "neural coding problem," its goal is to determine what features of neural activity carry information about external stimuli. This approach has led to many empirical and theoretical proposals about the spatial and temporal features of neural population activity, or "neural codes," that represent sensory information (Buonomano and Maass, 2009; Harvey et al., 2012, 2013; Kayser et al., 2009; Luczak et al., 2015; Panzeri et al., 2010; Shamir, 2014). However, there is still no consensus about the neural code for most sensory stimuli in most areas of the nervous system.

The lack of consensus arises in part because, while it is established that certain features of neural population responses carry information about specific stimuli, it is unclear whether the brain uses the information in these features to perform sensory perception (Engineer et al., 2008; Jacobs et al., 2009; Luna et al., 2005; Victor and Nirenberg, 2008). In principle, the link between sensory information that is present and sensory information that is read out to inform choices can be probed using the animal's behavioral report of sensory stimuli. In addition, improvements in techniques to perturb activity of neural populations during behavior (Boyden et al., 2005; Deisseroth and Schnitzer, 2013; Emiliani et al., 2015; Tehovnik et al., 2006) now make it possible to test causally hypotheses about the neural code, by "writing" on the neural tissue putative information and then measuring the behavior elicited by this manipulation. However, progress in cracking the neural code has been limited by the lack of a conceptual framework that fully integrates the advantages offered by behavioral, neurophysiological, statistical, and interventional techniques.

Here we elaborate such a conceptual framework, which at its core is based on a change in how a neural code should be defined. We propose that a neural code should be defined as the set of neural response features carrying sensory information that, crucially, is used by the animal to drive appropriate behavior; that is, the set of neural response features that have an intersection between sensory and choice information. In the following, we discuss this framework and its implications for designing and interpreting experiments aimed at cracking the neural code, as well as some theoretical and experimental challenges that arise from it.

What It Takes to Crack the Neural Code Underlying a Sensory Percept

To illustrate our new framework, we consider a perceptual discrimination task in which an animal has to extract information present in the sensory environment and, based on that information, choose an appropriate action. For definiteness, we assume (Figure 1A) a two-alternative forced-choice discrimination task: the animal has to extract color information from a visual stimulus (that is decide whether a green [$s = 1$] or a blue [$s = 2$] stimulus was presented) and choose accordingly to move left (choice $c = 1$) or right ($c = 2$), with the correct choice resulting in a reward (we numbered choices so that $c = 1$ is the correct rewarded choice for $s = 1$ and $c = 2$ is the correct choice for $s = 2$). We suppose that an experimenter is recording the activity of a

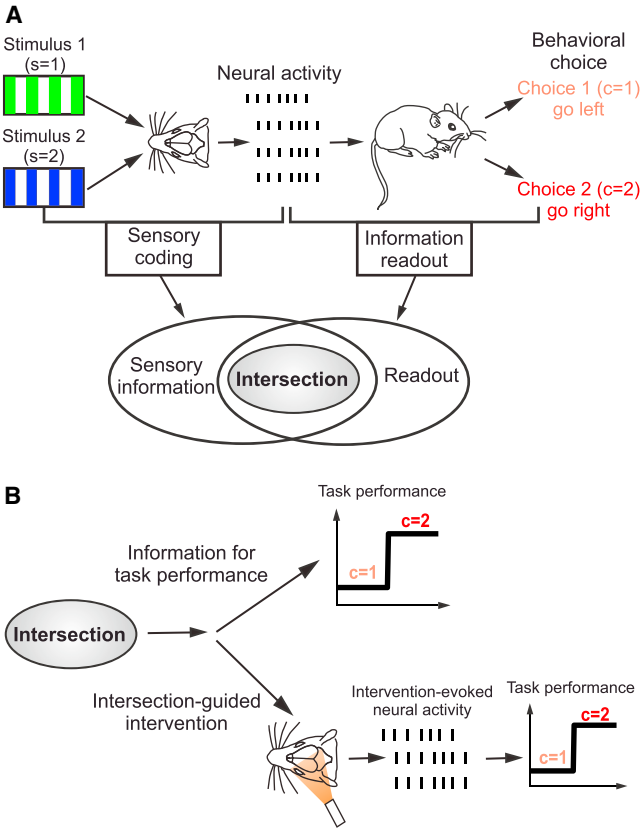


Figure 1. Intersection Information Helps Combining Statistics, Neural Recordings, Behavior, and Intervention to Crack the Neural Code for Sensory Perception

(A) Schematic showing two crucial stages in the information processing chain for sensory perception: sensory coding and information readout. In this example, an animal must discriminate between two stimuli of different color ($s = 1$, green; and $s = 2$, blue) and make an appropriate choice ($c = 1$, pink; and $c = 2$, red). Sensory coding expresses how different stimuli are encoded by different neural activity patterns. Information readout is the process by which information is extracted from single-trial neural population activity to inform behavioral choice. The intersection between sensory coding and information readout is defined as the features of neuronal activity that carry sensory information that is read out to inform a behavioral choice. Note that, as explained in the main text, a neural feature may show both sensory information and choice information but have no intersection information; this is visualized here by plotting the intersection information domain in the space of neural features as smaller than the overlap between the sensory coding and information readout domains.

(B) Only information at the intersection between sensory coding and readout contributes to task performance. Neural population response features that belong to this intersection can be identified by statistical analysis of neural recordings during behavior. Interventional (e.g., optogenetics) manipulations of neural activity informed by statistical analysis of sensory information coding can then be used to causally probe the contribution of neural features to task performance at this intersection.

population of sensory neurons (visual neurons in this example) while the animal performs the task. We would like to determine whether the activity of these neurons contributes causally to the animal's perception and behavioral choice.

The neural code in tasks such as this one involves two crucial stages in the information processing chain (Figure 1A). The first stage is *sensory coding*: the mapping, on each trial, of the sensory stimulus to neural population activity. The second stage is *information readout*: the mapping from neural population activity to behavioral choice.

An important observation is that sensory coding and information readout can be based on distinct features: some features of neural activity used to encode sensory information may be ignored when information is read out, and vice versa. For example, suppose that, as is often found (Panzeri et al., 2010; Shriki et al., 2012; Shusterman et al., 2011; Victor, 2000; Zuo et al., 2015), the precise timing of spikes and the spike rate are both informative about the stimulus. Although there is information in spike timing, the downstream neural circuit may be sensitive only to the rate, and thus unable to use information contained in spike timing. Or, the downstream circuit may be sensitive to spike timing, but if extracting spike timing information requires an independent knowledge of the stimulus presentation time to which the downstream circuit does not have access, the readout will not be able to use spike timing information. In a less extreme case, the same features of neural activity may be used for both sensory coding and information readout, but they may be weighted differently by different sets of neurons. For example, the timing of spikes may carry more information about the stimulus than the number of spikes, but the spike rate, defined as the spike count per unit time, may weigh more than spike timing in reaching a behavioral choice.

Although characterizing separately the features of neural activity used for sensory coding and for information readout can provide insight into neural information transmission, we argue that to crack the neural code it is essential to consider the *intersection* between sensory coding and information readout (Figure 1A), defined as the set of neural features carrying sensory information that is read out to inform a behavioral choice. The only information that matters for task performance is the information at this intersection. In fact, only features that lie at this intersection can be used to convert sensory perception into appropriate behavioral actions and can help the animal perform a perceptual discrimination task. We therefore define the neural code that allows the animal to do the task to be the "intersection" features of neural activity carrying sensory information that is read out for behavioral choice.

In the following, we propose a framework for identifying the information at the intersection of sensory coding and information readout. We propose a combination of statistical approaches, behavior, and interventional manipulations (Figure 1B). Statistical approaches can be used on single trials to identify the neural activity features that covary with the sensory stimuli and behavioral choices; they are, therefore, critical for forming hypotheses about the features of the neural activity that both contain sensory information and are used by the information readout. These hypotheses can be tested using experiments in which sensory stimuli are replaced with (or accompanied by) direct manipulation of neural population activity (Figure 1B). The manipulation of the specific features of neural population activity that take part in sensory coding and the examination of how these manipulations affect the animal's behavioral choices probe causally the intersection between sensory information and readout.

Examples of Candidate Neural Codes

Before detailing the concepts behind this proposed framework, we first provide examples to illustrate the types of neural codes

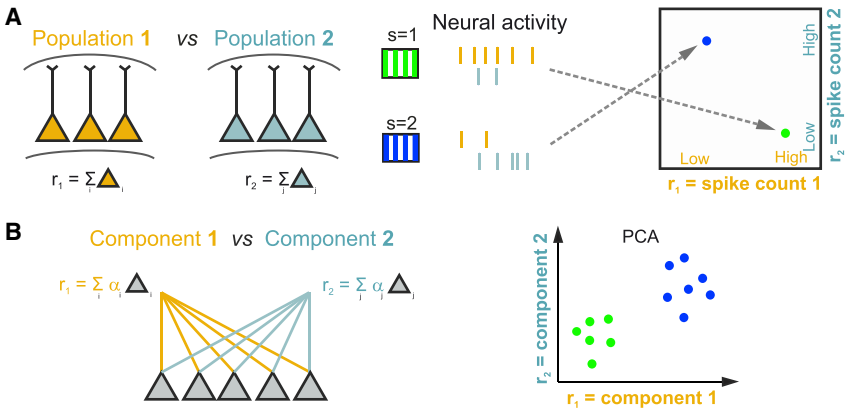


Figure 2. Schematic of Possible Pairs of Neural Population Features Involved in Sensory Perception

(A) Features r_1 and r_2 are the pooled firing rates of two neuronal populations (yellow and cyan) that encode two different visual stimuli ($s = 1$, green; and $s = 2$, blue). Values of single-trial responses of each population can be represented as dots in the two-dimensional plot of spike count variables in the r_1,r_2 space (rightmost panel in A).

(B) Features r_1 and r_2 are low-dimensional projections of large-population activity (computed for example with PCA as weighted sum of the activity of the neurons).

(C) Features r_1 and r_2 are spike timing and spike count of a neuron.

(D) Features r_1 and r_2 are the temporal regularity of the spike train of a neuron and spike count.

and questions that could be addressed. In all these examples, we suppose that we record (either from the same brain location or from multiple locations) neural population activity. That activity consists of n neural features, denoted $r_1,.., r_n$. We would like to determine which of these features, either individually or jointly, carry sensory information that is essential for performing the perceptual discrimination task. For simplicity, hereafter we focus on two features, r_1 and r_2, but our framework is general enough to deal with an arbitrary number of features (see Supplemental Information).

A common example of studies of population coding (Figure 2A) considers as candidate neural codes two features, r_1 and r_2, defined respectively as the total spike count of two populations of neurons. This has been the focus of many recent studies designed to test whether activity in specific neural populations is essential for accurate performance in sensory discrimination tasks (Chen et al., 2011; Guo et al., 2014; Hernández et al., 2010; Peng et al., 2015). Those spike counts could be from spatially separated populations in two different brain regions, as shown in Figure 2A, or they could be from different genetically or functionally defined cell classes in the same brain region (Baden et al., 2016; Chen et al., 2013; Li et al., 2015; Wilson et al., 2012). More sophisticated examples of features of neural population activity may involve low-dimensional projections of the activity of large neuronal populations (Cunningham and Yu, 2014). These could be, for example, the first two principal components of neural population activity and would consist of weighted sums of spike counts of the recorded population (Figure 2B). Major open questions that follow from these studies are (Otchy et al., 2015): which populations are instructive for the task (provide sensory information used for perceptual discrimination), which populations are permissive for the task (modulate task performance without directly contributing any specific sensory information), and which populations have no causal role in the sensory discrimination task despite having sensory information? The neural code is expected to be present in instructive populations. In contrast, permissive areas could provide task-relevant modulation that is not related to the sensory stimuli, such as attention or saliency signals. Populations with no causal role may still contain task-related information if it is inherited from instructive regions.

Other questions relevant for population coding regard which neurons are required for sensory information coding and perception (Houweling and Brecht, 2008; Huber et al., 2008; Reich et al., 2001). For example, often only a relatively small fraction of neurons in a population have sharp tuning profiles to the stimuli, whereas the majority of neurons have weak and/or mixed tuning to many different variables (Meister et al., 2013; Rigotti et al., 2013). Information about stimuli can be decoded from both types of neurons, but it remains a major open question whether only the sharply tuned neurons or other neurons as well can contribute to behavioral discrimination (Morcos and Harvey, 2016). A related question is: how many neurons are required for sensory perception? This question can be investigated by determining the smallest subpopulation of neurons that carries all information used for perception.

Another set of questions considers the role of spike timing in sensory coding and perception (Figures 2C and 2D). Spike timing could be measured with respect to the stimulus presentation time, an internal brain rhythm (Kayser et al., 2009; O'Keefe and Recce, 1993), or a rhythmic active sampling process such as sniffing (Shusterman et al., 2011). In many cases both spike timing and spike count carry sensory information (in the example of Figure 2C stimulus $s = 1$ elicits responses with fewer and earlier spikes than does $s = 2$). Although it is accepted that spike timing carries sensory information, whether or not timing is used for behavior has been vigorously debated (Engineer et al., 2008; Harvey et al., 2013; Jacobs et al., 2009; Luna et al., 2005; Zuo et al., 2015). For example, it is still debated whether the sensory information carried by millisecond-scale spike timing is redundant with that provided by the total spike count in a longer

response window of hundreds of milliseconds, whether the information in spike timing measured with respect to stimulus onset can be accessed by a downstream neural decoder, and whether recurrent circuits in higher cortical areas can extract millisecond-scale information.

Also of interest is whether the complex aspects of the temporal structure of spike trains could be part of the neural code. One possibility is that the regularity of spike timing of single neurons or the coordination of spike timing across cells carries information about the stimulus (as in the example of Figure 2D, where stimulus 1 elicits more regular spike trains than stimulus 2). The regularity or temporal coordination across cells may also have a large effect on the readout (Doron et al., 2014; Jia et al., 2013; Nikolić et al., 2013): for example, spikes closer in time may elicit a larger post-synaptic response and so may have a crucial impact on task performance. However, some studies have suggested that temporal coordination does not have a behavioral effect, but instead all spikes are weighted the same by the readout (Histed and Maunsell, 2014).

In what follows, we will consider, for simplicity, two features, and we will generically refer to these features as r_1 and r_2. These features could refer to spike timing and spike counts, the mean firing rate in two different brain regions, the activity of two different cell types, and so on.

Determining the Single-Trial Intersection between Sensory Information Coding and Information Readout Using Statistical Analysis of Neural Recordings

We now consider how to identify, using statistical measures applied to recordings of neural activity during behavior experiments, three conceptually important domains of interest in the neural information space (Figure 1A): the "sensory information" domain (the features of neural population activity that carry stimulus information), the "readout" domain (the features that influence the computation of choice), and the "intersection" between the two domains (the features that carry sensory information used to compute choice).

Throughout this Perspective, to illustrate neural coding and stimulus and choice domains, we use scatterplots of simulated responses characterized by two features; each dot in the two-dimensional feature plane (r_1,r_2) represents a single-trial response color coded for that trial's stimulus ($s = 1$: green; $s = 2$, blue) (see also Figure 2A, right panel for a schematization of this representation). Each dot therefore shows the simulated neural response of feature r_1 and r_2 on each individual trial.

A simple way to visualize how neural response features encode sensory stimuli is to compute a sensory decoding boundary (Quian Quiroga and Panzeri, 2009)—shortened to "*sensory boundary*" hereafter—that best separates trials by stimulus (i.e., that best separates the blue and green dots in the plots in Figure 3). This boundary (black dashed line in the r_1,r_2 plane in Figures 3A$_1$, 3B$_1$, and 3C$_1$) can be used as a rule to decide which stimulus most likely caused a given single-trial neural response. Similarly, we can visualize how neural response features are used to produce a choice with a "*decision boundary*" (Haefner et al., 2013), visualized as a red dashed line in Figures 3A$_1$, 3B$_1$, and 3C$_1$. This decision boundary is the line that best separates trials by choice, and in the specific simulated examples in Figure 3 it coincides with the actual boundary used to produce choice. Responses that lead to correct choices are shown as filled dots; those leading to incorrect choice are shown as open circles. The orientation of the boundaries determines the relative importance of each feature in sensory coding or choice: a diagonal boundary gives weight to both features, whereas a horizontal or vertical line gives weight only to r_2 or only to r_1, respectively.

To quantify how well each feature or set of features carries information about stimulus or choice, we use the fraction correct. In terms of the illustrations of Figure 3, the fraction of correctly decoded stimuli is the fraction of green or blue dots that fall on the correct side of the sensory boundary (below or above the sensory boundary for the green, $s = 1$, and blue, $s = 2$, stimulus, respectively). Other measures, such as those based on signal detection theory (Britten et al., 1996; Shadlen et al., 1996) or information theory (Quian Quiroga and Panzeri, 2009) can be used instead and are discussed in Supplemental Information. We use fraction correct primarily because it is simple and intuitive, but we could use any of the other measures without changing the basic framework. To emphasize the generality of our reasoning, hereafter we often refer to fraction correct as "information." If the fraction correct refers to the decoded stimulus, we call it "sensory information" or "stimulus information"; if it refers to decoded choice we call it "choice information."

We say that a neural response feature, r_i, carries sensory or choice information if the value of the presented stimulus or the animal's choice can be predicted from the single-trial values of this feature. Stimulus information and sensory boundaries are typically computed by presenting two or more different stimuli, and quantifying how well the stimulus-specific distributions of neural response features are separated by sensory boundaries (Quian Quiroga and Panzeri, 2009). Choice information has been typically computed separately from stimulus information (Britten et al., 1996), by evaluating decision boundaries from distributions of responses with no sensory signal or at fixed sensory stimulus (to eliminate spurious choice variations of neural response arising from their stimulus-related variations).

To understand the neural code associated with a particular task, it is relatively obvious that we need to consider both stimulus and choice information. If a response feature carries stimulus but not choice information, then the sensory information it carries isn't used for the task. If a response feature carries choice but not stimulus information, then although it may contribute to choice, or relay or execute the result of the decision making, it still cannot be used per se to increase task performance because it does not carry information about the sensory variable to be discriminated (Koulakov et al., 2005). However, a fact that has been underappreciated so far is that a neural feature can carry both sensory and choice information but still not contribute to task performance. This could happen, for example, when features carry both stimulus and choice information, but the rule used to encode sensory information is incompatible with the rule used to read them out.

We illustrate this in Figure 3A. Suppose that in this figure, r_1 and r_2 are the times of the first spike of two different neurons. These features are signal correlated (Averbeck et al., 2006); that is, both neurons spike earlier (corresponding to smaller values of both r_1 and r_2 in the scatterplot in Figure 3A) to the green stimulus

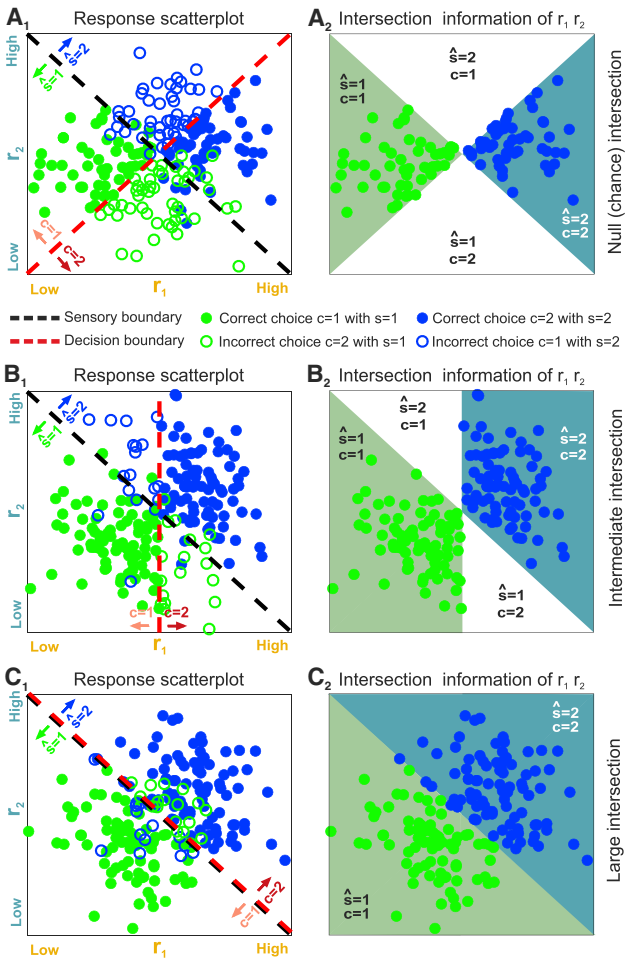


Figure 3. Impact of Response Features on Sensory Coding, Readout, and Intersection Information

In the left panels (A_1), (B_1), and (C_1), we illustrate stimulus and choice dependences of two hypothetical neural features, r_1 and r_2, with scatterplots of simulated neural responses to two stimuli, $s = 1$ or $s = 2$. The dots are color coded: green if $s = 1$ and blue if $s = 2$. 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. Panels (A_2), (B_2), and (C_2) plot only the trials that contribute to the calculation of intersection information. Those are the behaviorally correct trials (filled circles) in the two regions of the r_1, r_2 plane regions in which the decoded stimulus $\hat{s}$ and the behavioral choice are both correct. Each region is color coded with the color of the stimulus that contributes to it. White regions indicate the portion of the r_1,r_2 plane that cannot contribute to the intersection because for these responses either the decoded stimulus or choice is incorrect. The larger the colored areas and the number of dots included in panels (A_2), (B_2), and (C_2), the larger the intersection information. Panels (A_3), (B_3), and (C_3) plot a possible neural circuit diagram that could lead to the considered result. In these panels s indicates the sensory stimulus, r_i indicate the neural features and c the readout neural system, and arrows indicate directed information transfer: (A_1–A_3) no intersection information (the sensory and decision boundary are orthogonal); (B_1–B_3) intermediate intersection information (the sensory and decision boundary are partly aligned); (C_1–C_3) large intersection (the sensory and decision boundary are fully aligned).

($s = 1$) and later (corresponding to larger values of r_1 and r_2 in the scatterplot) to the blue stimulus ($s = 2$), with no "noise" correlations (Averbeck et al., 2006) between the activity of these neurons at fixed stimulus. For this encoding scheme, higher values of r_1+r_2 indicate that the blue stimulus is more likely, and so the sensory boundary is anti-diagonal: it is the line r_1+r_2 = constant. Suppose, though, that the readout does not have access to the stimulus time. In such a case, the only information the readout can use is the relative time of firing between the two neurons. This is the difference, r_1-r_2, and so the decision boundary is r_1-r_2 = constant. In this case, the responses carry information about both stimulus and the choice, but the responses cannot be used to perform the task – the orthogonal sensory and decision boundaries mean the animal's choice is unrelated to the stimulus.

The case illustrated in Figure 3A could happen also in studying neural population coding rather than spike timing. For example, r_1 and r_2 could be weighted sums of activity of neurons within a large population (as in Figure 2B), with stimulus encoded by the sum of the two neural features and choice by the most active of the two features (which feature is most active is revealed by the sign of r_1-r_2). Also in this "population" interpretation of Figure 3A, none of the stimulus information in population activity could be used to perform the task.

Investigating whether neural stimulus information is usefully read out for task performance requires quantifying whether neural discrimination predicts behavioral discrimination. This has traditionally been addressed by evaluating the similarity between neurometric functions (quantifying the trial-averaged performance in discriminating various pairs of stimuli using one or more neural response features) and psychometric functions (quantifying the animal's trial-averaged performance in discriminating the same set of pairs of stimuli). If a set of response features contributes to task performance, psychometric and neurometric functions should be similar (stimulus pairs discriminations that are easier for the animal should also be easier for the considered response features, and so on). This approach has provided numerous insights in sensory coding across several modalities (Engineer et al., 2008; Newsome et al., 1989; Romo and Salinas, 2003). For example, it was used to study the role of spike counts and spike times of somatosensory neurons for tactile perception of low-frequency (8–16 Hz) skin vibrations (Romo and Salinas, 2003). Although most such neurons encoded vibration frequency by spike count, some neurons encoded it by spike times fired in phase with skin deflections. However, the neurometric performance of spike counts correlated better to the psychometric one than that of spike times, suggesting that spike rates produce

this sensation (Romo and Salinas, 2003). A similar approach applied to high-frequency vibrations (>100 Hz) suggested that discriminating high-frequency vibrations relies on both spike times and counts (Harvey et al., 2013).

A potential problem with comparing neurometric and psychometric functions is that these functions may be similar even when the sensory and choice information do not intersect at all. The reason why this may happen is that it is based on comparing trial-averaged quantities, rather than comparing sensory information and animal's choice in single trials. To understand the possible problems of only comparing neurometric and psychometric functions, consider a new scenario (Figure S1A). The scenario is in part similar to that of Figure 3A: r_1, r_2 are again the first spike times of two different neurons, and they are tuned to the stimuli and contribute to choice; and, as in Figure 3A, in this new example of Figure S1A both neurons spike earlier to the green stimulus ($s = 1$) and later to the blue one ($s = 2$), leading to an anti-diagonal sensory boundary (line r_1+r_2 = constant), and the readout uses the relative time of firing r_1-r_2 between the two neurons (the decision boundary projected on the r_1,r_2 plane is r_1-r_2 = constant). However, suppose that in the example of Figure S1A the actual choice (unlike in Figure 3A) depends also on a third neural feature, r_3, which we'll take to be the sum of the spike count of the two neurons. Assume also that, crucially, the stimulus dependence of the spike count r_3 is similar to that of both r_1 and r_2, so that stimulus $s = 1$ elicits both earlier spike and lower counts than stimulus $s = 2$ does. Suppose finally that the experimenter now tunes the task difficulty by varying some "stimulus signal intensity" parameter whose effect on neural firing is to change the separation between the clouds of the $s = 1$ (green) and $s = 2$ (blue) stimulus-specific responses (Figure S1A$_2$). As the task becomes more difficult, the animal's psychometric performance decreases, as does the decoding neurometric performance (because the blue and green stimulus-specific distributions of points get closer). We can plot neurometric and psychometric performance as a function of signal intensity, and they will have similar shape: both will be near chance when signal intensity is small and the stimulus-specific distributions of r_1,r_2 largely overlap (Figure S1A$_3$), and will be nearly perfect when signal intensity is high and the stimulus-specific distributions of r_1,r_2 are far apart (Figure S1A$_5$). Thus, in this example (Figure S1A), statistical analysis will show that spike timing features r_1,r_2 have sensory information (because r_1+r_2 is stimulus dependent), have choice information (even at fixed stimulus, reflecting that r_1-r_2 impacts on choice), and the neurometric function of r_1,r_2 is similar to the psychometric function (because the stimulus dependence of both r_1 and r_2 is similar to that of the firing rate r_3, which is the only contributor to task performance). Yet r_1,r_2 do not contribute to task performance because none of the sensory information they carry is read due to the orthogonality of the sensory and decision boundaries. That r_1,r_2 do not contribute to task performance can only be discovered by observing that the trial-to-trial fluctuations of the accuracy of sensory information in r_1,r_2, encoded only by r_1+r_2, does not influence at all behavior, as the decision depends on r_1-r_2. For example, in trials when r_1+r_2 indicates the presence of a stimulus different from that presented, behavior is not less (or more) likely to be correct because of this stimulus coding error (Figures S1A$_3$–S1A$_5$).

These examples illustrate a general fact: it is not possible to determine whether sensory information is transmitted to the readout using the trial-averaged stimulus and choice information, either separately or in combination. It is, instead, necessary to investigate the effect of sensory coding on information readout within a single trial. We therefore propose the use of a measure we call intersection information, denoted *II*. Conceptually, intersection information is large only if the neural features carry a large amount of information about the stimulus and that information is used to inform choice—so that, based on these features, the animal is correct most of the time.

A quantitative description of *II* was derived in Zuo et al. (2015). The authors reasoned that, if feature r_i contributes to task performance, there should be an association on each trial between the accuracy of sensory information provided through that feature and behavioral choice. In other words, on trials in which r_i provides accurate sensory evidence (stimulus is decoded correctly from r_i), then the likelihood of correct choice should increase. Thus, the simplest operational definition of the intersection information, *II*, for a particular feature is the probability that on a single trial the stimulus is decoded correctly from r_i and the animal makes the correct choice (see Supplemental Information for additional details, in particular Eq. S7).

Intersection information can be used to rank features according to their potential importance for task performance. Importantly, it is high if there is a large amount of stimulus information and readout is near optimal. It is low, on the other hand, if a neural response feature has only sensory information but very little choice information, or vice versa, or if the rule used for sensory coding is incompatible with the rule used by the readout.

We illustrate intersection information using three examples (Figures 3A$_1$, 3B$_1$, and 3C$_1$), with null (chance-level), intermediate, and high values of intersection information, respectively. In these plots, we divide the r_1,r_2 feature space into four possible areas based on the sensory and decision boundaries: $\hat{s} = 1$, $c = 1$; $\hat{s} = 1$, $c = 2$; $\hat{s} = 2$, $c = 1$; $\hat{s} = 2$, $c = 2$ ($\hat{s}$ is the decoded stimulus, which can be different from the stimulus, s, presented to the animal). The intersection information is the fraction of trials that are decoded correctly and result in a correct behavioral choice; these trials correspond to the filled dots, indicating a trial with correct choice, shown in the regions in Figures 3A$_2$, 3B$_2$, and 3C$_2$ colored with the decoded stimulus color code. The larger are the colored areas, the larger is the intersection information. Chance level for the intersection measure is when there is no relationship between the stimulus decoded by neural activity and the choice taken by the animal at fixed stimulus (the chance level of intersection equals the product of the probability of a correct behavioral choice and the probability of correctly decoding the stimulus; see Supplemental Information for details). This is the case in Figure 3A, where the sensory and decision boundaries are orthogonal. Because trials that provide faithful stimulus information are just as likely to result in correct as incorrect choices, there is chance-level intersection information (see Figure S1D for the intersection information values in these examples).

Intersection information is intermediate when only some of the features of neural activity carrying sensory information are read out while the information of others is lost before the readout

stage. This is the case in Figure 3B, where both r_1 and r_2 carry sensory information but only r_2 is read out. This may correspond, for example, to a case when both spike count r_1 and spike timing r_2 of a neuron carry information, but only count r_1 is used for behavior (similarly to the case of O'Connor et al., 2013), for example, because the readout mechanism is not sensitive enough to precise spike timing.

Intersection information is largest when the optimal sensory boundary and the decision boundary coincide, as in Figure 3C, so that all sensory information is optimally used to perform the task. This is the case when all measured features of neural activity that carry sensory information directly contribute to the animal's choice. In this example (Figure 3C), trials that lead to correct stimulus decoding from the joint features, r_1 and r_2 (those below the diagonal for the green stimulus, $s = 1$, and those above the diagonal for the blue stimulus, $s = 2$), always lead to correct behavioral choices. Trials leading to incorrect stimulus decoding from r_1 and r_2 (above the diagonal for $s = 1$ and below it for $s = 2$) always lead to incorrect behavioral choices. This situation is reminiscent of texture encoding by somatosensory cortical neurons (Zuo et al., 2015), in which both spike rate and timing seem to carry sensory information that is used for behavioral discrimination.

The above simple reasoning can be extended to provide more refined measurements of the relationship between sensory information in neural activity and behavioral choice. For example, one could also measure (Zuo et al., 2015) what we call the fraction of intersection information *fII*, defined as the fraction of trials with correct stimulus decoding that have correct behavior. Unlike *II*, *fII* is not sensitive to the amount of sensory information (the fraction of trials the stimulus is decoded correctly from neural feature *r*), but only to the proportion of these correctly decoded trials that lead to correct behavior. Thus, *fII* is an indicator of the optimality of the readout—in the linear case, the alignment between the sensory and decision boundaries—rather than the total impact of the code on task performance. Measuring both *II* and *fII* could be useful to determine whether a moderate amount of intersection information, *II*, is because the feature has a moderate amount of information but is efficiently read or because the feature has high information but not read out very efficiently. Moreover, given that if a feature r_i contributes to task performance, then, in trials when r_i provides inaccurate evidence (stimulus is decoded incorrectly), the likelihood of correct choice should decrease, an additional separate quantification of the agreement of stimulus information and behavioral choice in incorrect trials would complement intersection measures (see Zuo et al., 2015 and Supplemental Information).

The purely statistical approach to measure intersection information is most straightforward if all response features are statistically independent, because in that case the intersection information approach applied to a set of features would unambiguously identify the contribution of those features to task performance. However, often features are not independent. For example, if the features are the activity of neurons in different brain regions, these features might be partly correlated if there are connections between the two regions. Alternatively, if the features are spike timing and spike count, both involve the same spikes and so may be dependent. The presence of dependencies among features complicates the interpretation of intersection information. In particular, it raises two critical questions. First, does a set of features with intersection information contribute to task performance, or instead reflect only a correlation with other features that truly contributes to task performance? Second, does each neural response feature provide unique intersection information that is not provided by other features?

To illustrate the complications induced by correlations among features, we consider the intersection information from one feature rather than two. We return to Figure 3, for which responses are signal correlated in all panels (that is, responses to $s = 1$ are on average lower than those to $s = 2$ for both features [Averbeck et al., 2006]). We first consider a case (as in Figure 3B) for which both features carry information about the stimulus and are partly correlated (because of signal correlations) but only feature r_1 is read out. Suppose that we apply our statistical analysis to feature r_2; that is, we decode the stimulus using only r_2, which can be done optimally by decoding responses in the lower and upper half of the r_1,r_2 space as $\hat{s} = 1$ and $\hat{s} = 2$, respectively. We will find higher-than-chance intersection and choice information (as shown by the fact that lower values of r_2 are found in trials with choice $c = 1$ than in trials with $c = 2$, see Figure S1B) even though r_2 is not read out. That's only because r_2 is correlated with r_1, which is the feature that is truly read out.

If we record from both features, we can differentiate, just from statistical analyses of neural recordings, between the case when only one feature is read out (Figure 3B) and the case when both are read out (Figure 3C). If, as in Figure 3C, r_1 and r_2 carry complementary sensory information (the diagonal sensory boundary implies that both features should be used for optimal decoding) and if the readout uses both features (the decision boundary is also diagonal), then intersection information obtained when decoding the stimulus using two features will be larger than the intersection information obtained when decoding the stimulus with either feature alone. This is because decoding the stimulus with only one feature will lose the complementary task information present in the other feature and so the task performance will suffer (Figures S1C and S1D). Thus, a statistical signature that task performance benefits from both features is that using both feature increases the intersection information (Figure S1D). However, if we cannot record both features (and, more generally, all features that carry intersection information), the only way to fully prove which features contribute to task performance is to use interventional methods. That's the subject of the next section.

Causal Interventional Testing of the Neural Code

Why Do We Need Intervention?

The statistical methods described above for determining sensory, choice, and intersection information are useful for identifying potential neural coding mechanisms, and for forming hypotheses about information coding and transmission. However, as just discussed, because response features are often correlated and because we do not usually have experimental access to all of them, whether a neural feature carries information in the intersection between sensory coding and readout can ultimately only be proved with intervention. Before discussing how to

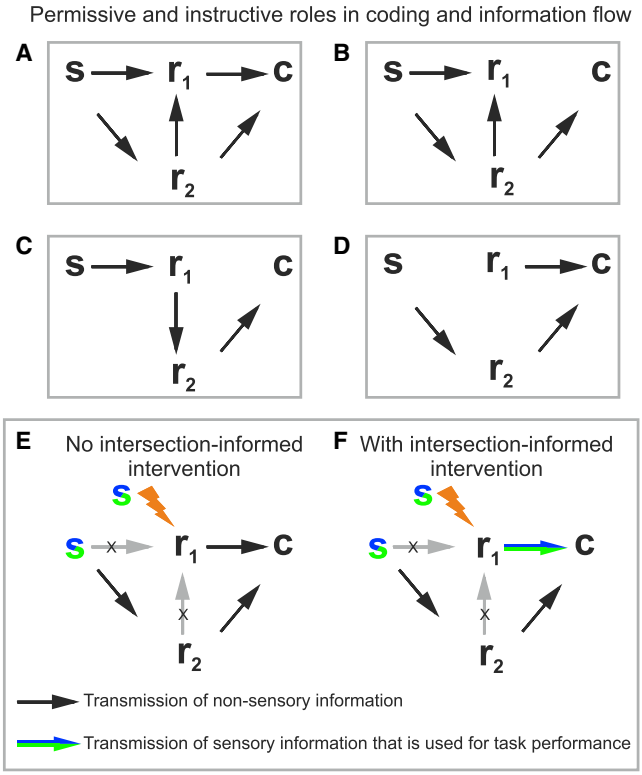


Figure 4. Causal Manipulations to Study the Permissive and Instructive Roles in Coding and Information Flow

(A–D) Interventional approaches can be used to disambiguate among different conditions. (A) The neural features r_1 and r_2 carry significant information about the stimulus, s, and provide essential stimulus information to the decision readout, c. (B) r_1 does not send information to c, but only receives a copy of the information via r_2, which does send stimulus information to c. (C) r_1 provides instructive information about s to r_2 and r_2 informs c instructively; (D) r_1 influences c but does not directly carry information about s.

(E and F) Interventional approaches can be used to reveal cases in which r_1 informs c but does not send stimulus information that contributes to task performance (black arrow in E) from cases in which r_1 sends stimulus information used for decisions (colored arrow in F).

design interventional experiments that test intersection information, it is useful to consider why interventional manipulations of neural activity are so crucial to prove hypotheses. (In the following, we refer to "statistical" information measures as shorthand for measures of information obtained from recorded natural unperturbed neural activity, and "interventional" information measures to indicate information estimates from neural activity imposed by intervention.)

Suppose that statistical measures like those described in the previous section found that a neural feature, r_1, carries stimulus, choice, and intersection information. An interpretation of this result is that r_1 provides essential stimulus information to the decision readout (this is indicated in Figure 4A by the arrow from r_1 to the choice, c). However, another interpretation (sketched in Figure 4B), one that is still compatible with these statistical measures, is that r_1 *does not* transmit information to c (not even indirectly). Instead, it only receives a copy of the information that other neural features (such as r_2 in Figure 4B) do transmit to the choice. In this case, the sensory information in r_1 is not causally involved in the decision (as indicated by the lack of arrow from r_1 to c in Figure 4B), but r_1 correlates with the decision because it correlates with r_2, the decision's cause.

Intervention can disambiguate these two scenarios by imposing a chosen value on r_1, one that is decided by the experimenter and so is independent of r_2. By doing that, we break any possible effect of r_2, or of any other possible variable, on r_1 (Figure 4E). In this case, any observed relationship between r_1 and choice must be due to the causal effect of r_1 on choice (Pearl, 2009).

In the following we discuss how to design a causal intervention experiment that tests whether neural features carry intersection information. We are interested in an intervention design that can tell whether r_1 transmits stimulus information used for decision (as in Figure 4F, indicated by the arrow between r_1 and choice c being colored like a stimulus) or r_1 informs c but does not transmit stimulus information contributing to task performance (as in Figure 4E, indicated by the arrow between r_1 and c not being colored).

Intervention on Neural Activity and Intersection between Sensory Information and Readout

Here we examine cases in which we can both record and manipulate (in the same animal, but not necessarily at the same time) neural features r_1 and r_2 during a perceptual discrimination task.

Let us first consider a causal intervention on the neural features. Suppose that we impose a number of different values of r_1, r_2 in a series of intervention trials ("lightning bolt" symbols in Figure 5, colored by the behavioral choice they elicit) and we measure the choice taken by the animal. In our examples, choice is determined by the red dashed decision boundary in the r_1, r_2 space. Observing the correspondence between the value imposed on r_1, r_2 and the animal's choice would easily determine the orientation of the decision boundary (Figures 5A–5C). From this interventionally determined decision boundary, choice information can be obtained exactly as in the statistical case.

Applying the same reasoning used for the statistical case, an interventional measure of intersection information is the fraction of trials on which the animal's choice reports the stimulus that would be decoded (using the sensory boundary acquired with statistical analysis of neural responses) from the imposed neural activity pattern (as above, this can be assessed against chance level). Application of this interventional measure of intersection information to our examples in Figure 5 shows that interventional intersection information captures the alignment of the sensory and decision boundary. It is high when, as in Figure 5C, the animal's choice ($c = 1$, pink; $c = 2$, dark red) always corresponds to the stimulus decoded from neural activity (in Figure 5C, the case of maximal intersection, all patterns in the $\hat{s} = 1$ "green" decoding region lead to $c = 1$, and the same applies to the $\hat{s} = 2$, $c = 2$ region); it is null (chance-level) when sensory and decision boundaries are mismatched (as in Figure 5A, where half of imposed patterns in either stimulus decoding region lead to choice $c = 1$ and half to $c = 2$).

A critical observation is that the intersection information computed via intervention may be different from that computed using purely statistical analysis. That can happen, as discussed above, if the neural features are correlated with variables that did

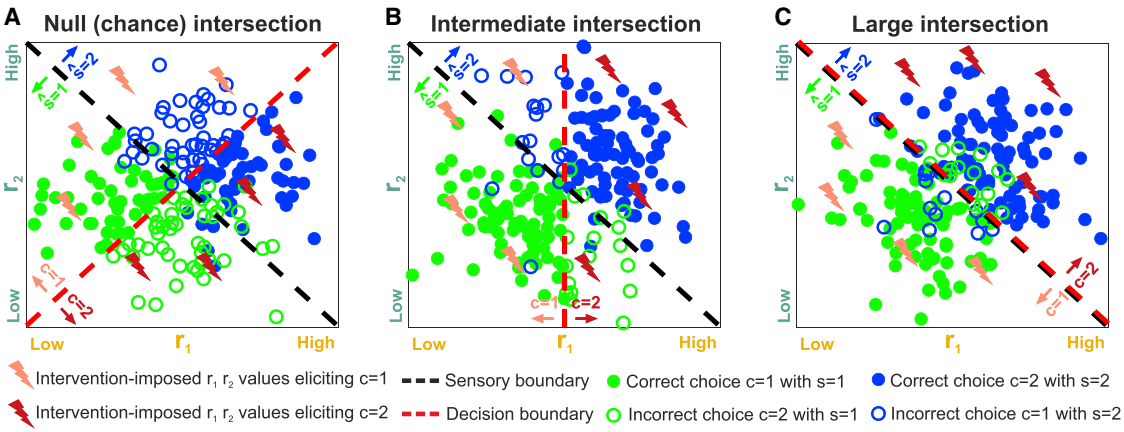


Figure 5. Schematic of an Experimental Design to Probe Intersection Information with Intervention

Three examples of neural responses (quantified by features r_1,r_2) to two stimuli, with conventions as in Figure 3. We assume that some patterns of neural activity are evoked by interventional manipulation in some other trials. The "lightning bolts" indicate activity patterns in r_1,r_2 space evoked by intervention: they are color coded with the choice that they elicited (as determined by the decision boundary—the dashed red line). Choice $c = 1$ is color coded as pink, and $c = 2$ as dark red. The choices evoked by the intervention can be used to determine, in a causal manner, the position of the decision boundary (as the line separating different choices). The correspondence between the stimulus that would be decoded from the neural responses to the intervention-induced choice can be used to compute interventional intersection information.

(A) A case with no interventional intersection information (the sensory and decision boundary are orthogonal).

(B) A case with intermediate intersection (the sensory and decision boundary are partly aligned).

(C) A case with large intersection (the sensory and decision boundary are fully aligned).

carry intersection information, but did not themselves provide any information about choice. For example, if a statistically determined non null choice or intersection information in one feature just reflects a top-down choice signal and not a causal contribution of the feature to choice, this feature will show null (chance-level) intersection information with intervention. Thus, of particular interest are cases in which the decision boundary is orthogonal to the sensory boundary under intervention, but not in trials without an intervention (O'Connor et al., 2013). When that's the case, the neural features under investigation carry no intersection information and the intersection or choice information determined statistically mean that the considered features only correlate with the true factors that are instructive for task performance.

The values of the interventionally evoked neural features in an experiment are arbitrarily determined by the experimenter. The chosen evoked neural features may be designed to drive behavior robustly but may occur very rarely during perception in natural conditions. This may lead to an over-estimation of their importance for task performance. To correct for this problem, when computing interventional intersection information, we should weigh intervention results with the probability distribution over stimuli and responses that occur under natural conditions (see Supplemental Information). Thus, evaluating the causal impact of a neural code with intervention experiments ultimately demands a statistical analysis of the probability of naturally occurring patterns during the presentation of each stimulus while performing the task.

By analogy to what we proposed for the statistical measures, we can design intervention experiments that address whether two neural response features, r_1, r_2, that are correlated during measures of natural neural activity both contribute causally to choice and to task performance. Suppose that (as in Figures 5B and 5C) we recorded two correlated features in unperturbed (i.e., no intervention) conditions and that we would like to determine interventionally whether the readout uses both such sources of sensory information to perform the task. Designing such an experiment requires manipulating both features at the same time, and then comparing interventional intersection information of the joint features and of the individual ones. If the experimenter designs a set of intervention patterns that generate uncorrelated feature values, then only the features that carry sensory information and are read out will show higher-than-chance intersection information. If the experimental design cannot fully decorrelate the features evoked by intervention, then a complementary contribution to task performance of the two features will still be revealed interventionally when finding that adding a feature increases the interventional intersection information, exactly as in the statistical analysis case.

Above we argued that statistical analysis is not sufficient to determine whether there is intersection information in a set of neural features. In the following, we argue that causal manipulations of neural activity alone are also not sufficient to determine whether there is intersection information in a set of neural features. Experiments are frequently designed such that an animal is trained to discriminate neural activity patterns that are created artificially using interventional approaches, such as microstimulation or optogenetics, without direct regard to how these patterns may encode sensory stimuli. This approach is extremely powerful for testing the capabilities of the readout. For example, this approach has been used to test the sensitivity of the readout to precisely timed neural activity (Doron et al., 2014; Yang et al., 2008; Yang and Zador, 2012) and also to test the minimal number of neurons or spikes to which a readout could be sensitive (Houweling and Brecht, 2008; Huber et al., 2008). Thus, this approach can be used to infer intersection information indirectly

(by comparing the features that carry sensory information with those that can be detected by the readout). However, this approach is insufficient to determine directly whether a given neural feature is used for performing specific sensory discrimination tasks, and to evaluate how much this feature contributes to sensory discrimination.

Because both statistical and interventional approaches by themselves are not sufficient to test neural codes, we propose a scheme in which statistical analyses must be used to generate hypotheses about neural codes, and interventional experiments must be used to test them.

Neurophysiological Examples of the Potential Advantages of Application of the Statistical and Interventional Concept of Intersection

The foundations underlying intersection between sensory information and readout can be traced back to the work of Newsome and colleagues on visual motion perception in primates (Britten et al., 1996; Newsome et al., 1989). These studies showed that visual area MT encoded visual motion information in its firing rate: higher firing rates indicate motion along the neuron's preferred direction. They established a statistical relationship between the animal's choice in a visual motion discrimination task and the firing rate of MT neurons in the same trial (Britten et al., 1996). The causal role of the firing rate of MT neurons in motion perception was interventionally demonstrated showing that microstimulation of this region biases perception of motion direction (Newsome et al., 1989). Such studies continue today, taking also advantage of modern genetic, optogenetic, and recording techniques. One example is the study of the neural coding of sweet and bitter taste in mice. The authors first established that anatomically separate populations of neurons responded to sweet and bitter taste, and thus carried stimulus information (Chen et al., 2011). An optogenetic intervention was then used to activate the spatially separated "sweet" and "bitter" populations (Peng et al., 2015). These intervention experiments elicited behavioral responses as expected for a mouse's response to sweet and bitter tastes. These studies therefore reveal intersection information in neural codes as spatially segregated response patterns using a combination of stimulus information, statistical analysis, and intervention.

These studies investigated simple properties of an individual neural feature (firing rate of classes of neurons) and followed implicitly part of the logic of the framework proposed here, although they did not measure a single-trial statistical intersection that we propose. Measuring the intersection information becomes, however, crucial in more complex scenarios in which (unlike the cases considered above) either a clear hypothesis about the neural code does not exist a priori (as may happen when analyzing coding of complex natural stimuli, rather than simpler laboratory stimuli) or when there are multiple, perhaps partly correlated, candidate features for the neural code that all seem, statistically, to contribute to choice or stimulus. In these cases, it is necessary to evaluate quantitatively the contribution of each feature to behavior. Below we discuss how the full or partial application of the ideas of our intersection framework in these more complex scenarios could provide further insight into the neural code.

The statistical intersection information framework has been applied to investigate whether millisecond-scale spike timing of somatosensory cortical neurons provides information that is used for performing a whisker-based texture discrimination task (Figures 6A and 6B), above that already carried by spike counts over timescales of tens of milliseconds (Zuo et al., 2015). The authors computed a spike-timing feature by projecting the single-trial spike train onto a timing template (constructed for optimal sensory discrimination) whose shape indicated the weight assigned to each spike depending on its timing (Figure 6C). Computed spike counts corresponded to weighting the spikes with a flat template, which assigns the same weight to spikes independent of their time. This provided timing and count features that had negligible correlation (the temporal distribution of spikes was largely independent from their total number). Both timing and count carried significant sensory (Figure 6D), choice (Figure 6E), and intersection (Figure 6F) information, with timing carrying more information than count for all these types of information. The joint information was, for all types of information, larger than the one carried by either feature alone. These results indicate that in this task sensory information was complementarily multiplexed in spike counts and timing and was also complementarily combined to perform the task. Of the two features, however, timing carried both more sensory information and had a greater influence on the animal's choice. Thus, the statistical intersection framework helped form a very precise hypothesis that multiplexing spike timing and spike count information is the key neural code used to solve the task. A further application of the interventional intersection framework, not yet applied to this experiment, would strongly prove or disprove this multiplexing hypothesis for texture coding.

This example illustrates that the statistical analysis of information intersection may be critical to correctly interpret the results of an interventional experiment and to refine its design. In this case, profound texture-dependent spike timing differences were found even across nearby neurons (Zuo et al., 2015). The cellular-level and millisecond-scale temporal resolution of this information coding revealed by the statistical analysis strongly constrains the interventional experimental design, as it indicates that finely spatially patterned and temporally precise intervention must be used to test whether spike timing is part of the neural code. Also, this example shows how statistical intersection results are essential to interpret successes and failures of interventions. For example, in the presence of such profound neuron-to-neuron differences in spike timing responses to textures, a causal effect of spike timing on behavior would not have been detected using a wide-field optogenetic intervention that activated all neurons simultaneously (see also section Considerations of Interventional Experimental Design). Statistical analysis would be essential to reveal that this failure would not have been because spike timing was not part of the neural code used to perform the task, but because the optogenetically induced activity did not preserve the natural texture-dependent timing differences across neurons.

A study (O'Connor et al., 2013) that implemented an approach close in spirit to the intersection information framework both at the statistical-analysis and interventional level is a recent investigation of the role of spike timing and spike rate coding in

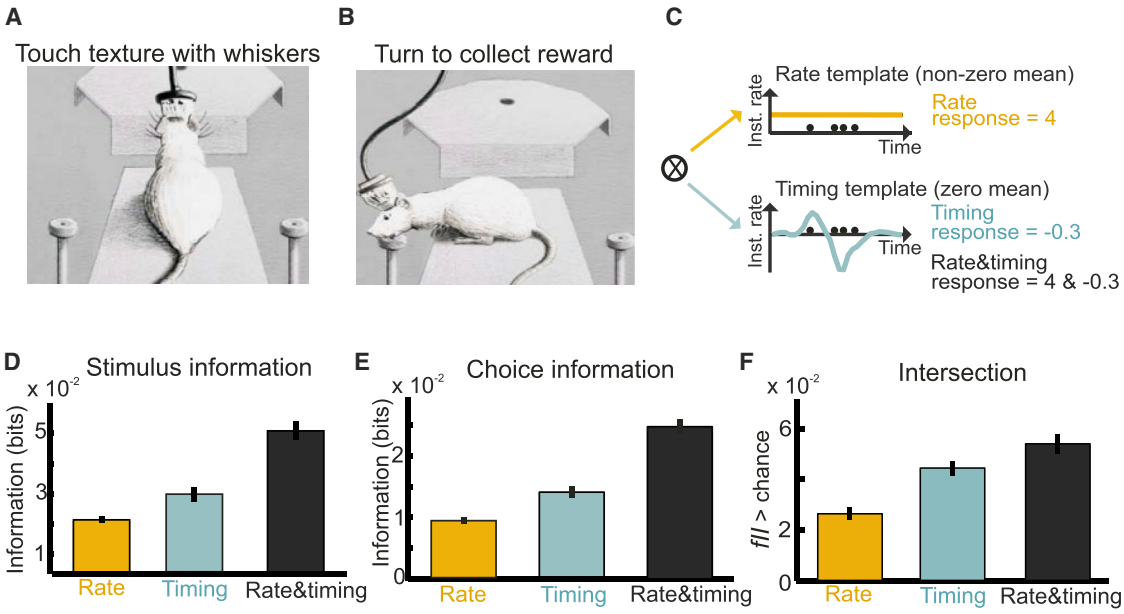


Figure 6. Examples of Statistical Intersection Measures in a Texture Discrimination Task

This figure shows how spike timing and spike count in primary somatosensory cortex encode textures of objects, and how this information contributes to whisker-based texture discrimination.

(A and B) Schematic of the texture discrimination task. (A) On each trial, the rat perched on the edge of the platform and extended to touch the texture with its whiskers. (B) Once the animal identified the texture, it turned to the left or the right drinking spout, where it collected the water reward.

(C) Schematic of the computation of spike count and spike timing signals in single trials.

(D–F) The mean ± SEM (over n = 459 units recorded in rat primary somatosensory cortex) of texture information (D), choice information (E), and fraction of intersection information *fII* (F). Modified with permission from Zuo et al. (2015).

whisker-based detection task of object location (Figure 7A). The authors found, based on statistical measures, that both timing and rate carried both stimulus and choice information (Figure 7B). The authors then probed the role of timing and rate by replacing the somatosensory object with optogenetic manipulation of layer 4 somatosensory neurons (Figures 7C and 7D). The authors found that using optogenetics to induce neural activity with information in rate caused the animal to report the sensation of a "virtual pole" (Figure 7C), whereas adding to this optogenetic manipulation information in spike timing relative to whisking did not elicit additional behavioral performance in virtual sensation (Figure 7D). When interpreted within our framework, these results suggest that spike times do not carry any intersection information that is additional to that carried by rates. An additional application of the statistical intersection framework to these neurophysiological recordings—not performed in that study—would allow a more precise evaluation of the impact of timing and rate codes on task performance (see previous section) and could provide an important independent confirmation of this hypothesis based on naturally evoked neural activity only.

Considerations of Interventional Experimental Design

Interventional approaches may involve use of one or more experimental techniques such as optogenetic (Lerner et al., 2016) and chemogenetic (Sternson and Roth, 2014) manipulations, intraparenchymal electrical stimulation (Tehovnik et al., 2006), transcranial direct current stimulation, and transcranial magnetic stimulation (Woods et al., 2016), to name a few. Given its unique combination of high cell-type specificity and temporal resolution, below we focus mostly on optogenetics.

There are at least two dimensions over which experimental design may be varied. One is how intervention is coupled with sensory stimuli; the other is how intervention is performed. In the following, we consider how these possible experimental variations along these dimensions relate to the intersection framework.

Virtual Sensation Interventional Experiments versus Experiments Overriding or Biasing Natural Sensory Signals

Our framework assumes that we test sensory encoding and information readout using a perceptual discrimination task. An important experimental design question is how to incorporate interventional approaches. Our focus is on understanding the codes that arise from natural sensory cues, and so we mainly consider cases in which interventional trials are interleaved with non-interventional ones.

One practical question for experimental design is whether on intervention trials the sensory stimulus should also be presented, or if the intervention manipulation should be applied in isolation. One possibility is a "virtual sensation" experiment (Figure 8A), in which patterns of neural activity are imposed by intervention in the absence of the sensory stimulus and the animal is asked to report the perception of one of the two sensory stimuli. A classic example is the work of Romo and colleagues (Romo et al., 1998; Romo and Salinas, 2003) demonstrating that cortical microstimulation can entirely substitute for tactile stimulation in a

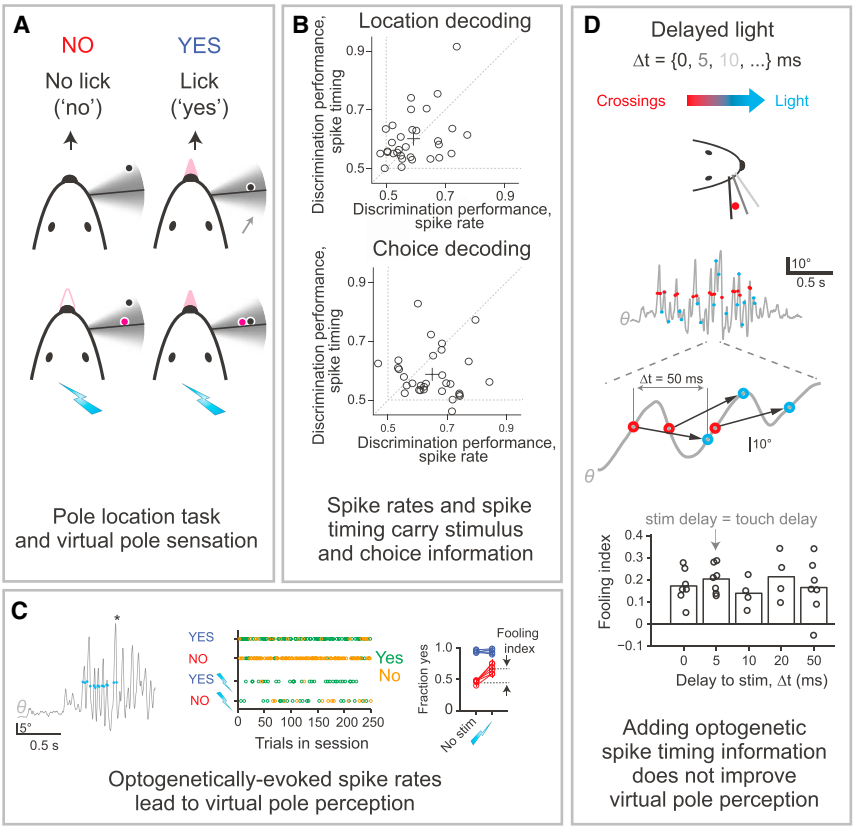


Figure 7. Examples of Statistical and Interventional Intersection Measures with Sensory and Illusory Touches

This figure shows results of the statistical and interventional test of the role of cortical spike timing and spike count in the neural code for whisker-based object location. The test involved closed-loop optogenetic stimulation causing illusory perception of object location.

(A) Schematic of the task: four trial types during a closed-loop optogenetic stimulation behavior session depending on pole location and optogenetic stimulation (cyan lightning bolts). A "virtual pole" (magenta) was located within the whisking range (gray area). Mice reported object location by licking or not licking.

(B) Decoding object location and behavioral choice from electrophysiologically recorded spikes in layer 4 of somatosensory cortex. Each dot corresponds to the decoding performance (fraction correct) of one neuron.

(C) Optogenetically imposed spike rates evoked virtual pole sensation. Left: optogenetic stimulation (blue circles) coupled to whisker movement (gray, whisking angle θ) during object location discrimination. Asterisk, answer lick. Middle: responses in the four trial types across one behavioral session. Green, yes responses; gold, no responses. Right: optogenetic stimulation in NO trials (red), but not in YES trials (blue), in barrel cortex increases the fraction of yes responses. Lightning bolt and "no stim" labels indicate the presence and absence of optogenetic stimulation, respectively. Error bars, SEM. Each line represents an individual animal.

(D) Adding timing information in the optogenetically evoked activity did not improve virtual pole perception. Top: delayed optogenetic stimulation was triggered by whisker crossing with variable delays, Δt. Middle: whisker movements with whisker crossing (red circles) and corresponding optogenetic stimuli (cyan circles) for Δt = 50 ms. Bottom: fooling index (fraction of trials reporting sensing of a virtual pole) as function of Δt. Modified with permission from O'Connor et al. (2013).

frequency discrimination task. Another example of virtual sensation is the induction of an illusory sensation of pole touching during whisking using optogenetic stimulation of cortical primary somatosensory neurons, as discussed above (see Figure 7 and O'Connor et al., 2013). The virtual sensation paradigm is very appealing because it can demonstrate the sufficiency of the considered neural code for creating sensation and for its direct relevance for the development of neural prosthetics.

Another possibility is to impose patterns in the presence of a sensory stimulus. This approach tests whether the imposed pattern can "override" or "bias" (Figure 8B) the signal from the sensory stimulus. A classic example of this approach can be found in the work of Newsome and colleagues (Salzman et al., 1990) showing that MT microstimulation in a visual motion discrimination task can bias the animal's perception toward the motion direction preferred by the neurons that were activated by microstimulation. A more recent example can be found in the study (also described above) examining the codes for sweet and bitter/salt taste sensation (Peng et al., 2015), where the authors showed that optogenetic activation of the sweet cortical field triggered fictive sweet sensation even in the presence of a salt stimulus. From the point of view of the formalism presented here, successfully overriding the signal from an opposite external stimulus is an appealing proof that the considered neural code provides information that is so crucial to the task that it can even win over other contrasting sources of information, such as those that may come from different or parallel pathways conveying information from the sensory periphery that contradicts the one injected through intervention of neural activity.

Considerations on How to Perform Intervention on Neural Activity, and the Advantages of Patterned Optogenetics

Imposing a pattern can be done in two conceptually distinct ways. In one, the experimenter mainly tries to "bias" (Guo et al., 2014; Li et al., 2015; O'Connor et al., 2013) the neural activity (Figure 8B). This consists of shifting the endogenous activity in a certain direction (for example, lowering the firing rates of a neural population by imposing a slight hyperpolarization or by exciting a set of inhibitory neurons). This can be done, for example, using wide-field, single-photon optogenetic stimulation of a network of a number of opsin-expressing neurons (this is illustrated in Figure 8C, note that the number of neurons in that sketch is limited to seven for presentation purposes only). A problem with this approach is that it does not completely remove correlations of the patterns evoked by intervention with other brain variables that are present in the endogenous component of the activity (because the evoked activity adds to the endogenous one). This means that this intervention may not entirely break the correlations among features or between

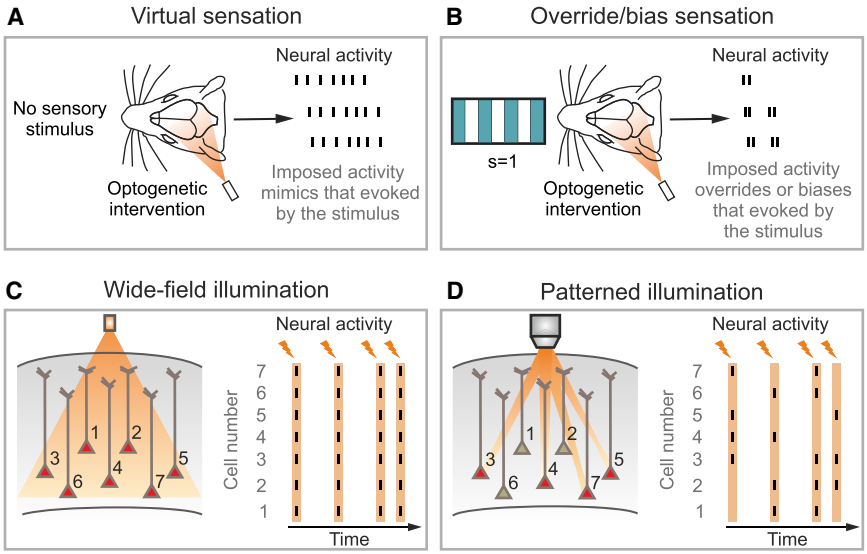


Figure 8. Experimental Configurations for Interventional Optogenetic Approaches

(A) In a virtual sensation experiment, the animal behavior is tested by applying the optogenetic intervention in the absence of the external sensory stimulus.

(B) Alternatively, optogenetic intervention can be paired with sensory stimulation with the aim of overriding or biasing neural activity evoked by the sensory stimulus.

(C) In the wide-field configuration for optogenetic manipulation, light is delivered with no spatial specificity within the illuminated area, resulting in the activation (red cells) of most opsin-positive neurons. Stimulation in this regime may lead to over-synchronous neural responses (right). The orange lightning bolts in the right panel indicate the time at which successive stimuli are applied. The neurons displayed in (C) and (D) are meant to represent a population of *N* neurons expressing the opsins; their number is here limited to 7 for presentation purposes only.

(D) Patterned illumination permits the delivery of photons precisely in space. When multiple and diverse light patterns are consecutively delivered (orange lightning bolts), optical activation of neural networks with complex spatial and temporal patterns becomes possible (right).

features and non-observed endogenous brain activity that the causal manipulations aim to remove. This is a concern particularly when investigating whether intersection information is complementarily carried by more than one feature, as an interventional bias may affect all features in a correlated way. For example, a general hyperpolarization of the population may both lower the spike rate and delay the latency of neural activity. Given the highly synchronous generation of photocurrents in opsin-expressing cells, wide-field optogenetics may even induce artificial correlations (Figure 8C).

The second interventional approach is to try to impose, or "write down" (Peron and Svoboda, 2011), a target neural activity pattern on a neural population (Figure 8D). This approach is, in principle, ideally suited to test hypotheses about the neural code, because it explicitly aims to overwrite endogenous activity, and so break down all sources of correlation. To crack a neural code, though, it needs to achieve high spatial and temporal precision. Recent optical developments (Bovetti and Fellin, 2015; Emiliani et al., 2015; Grosenick et al., 2015), termed patterned illumination, can deliver light to precise spatial locations (Figure 8D, see also Supplemental Information). When combined with light-sensitive optogenetic actuators, patterned illumination can perturb electrical activity with near cellular resolution (Baker et al., 2016; Carrillo-Reid et al., 2016; Packer et al., 2015; Papagiakoumou et al., 2010; Rickgauer et al., 2014).

Taking full advantage of the intersection framework will depend crucially on further development of improved optogenetics methods to "write" neural activity patterns. Current technologies target simultaneously a few dozen cells with a temporal resolution of few milliseconds (Emiliani et al., 2015). Major areas of future developments include scaling up of the number of stimulated neurons while maintaining single-cell resolution, improving temporal resolution, performing large-scale 3D stimulation, and precisely quantifying tissue photodamage during intervention. In addition, it will, ultimately, be important to implement these technologies with a closed-loop system (Grosenick et al., 2015), so that intervention can be tied to behavior. This will be useful, among other things, to predict and discount residual effects of endogenous activity (Ahmadian et al., 2011). In fact, both the number of responsive neurons and their functional responses to the sensory stimulus and to the intervention may vary as a function of behavioral variables such as arousal, attention, or locomotion that are reflected in brain states and ongoing neural activity (see also next section Potential Confounds). Coupling functional imaging with optogenetic intervention allows tracking these changes and adapting patterned photostimulation to brain dynamics. Moreover, because patterned illumination requires knowing where the cells to stimulate are, and what pattern to stimulate them with, it will be necessary to combine imaging with patterned photostimulation. Finally, taking full advantage of the intersection approach will require multimodal recording techniques. While electrophysiological recordings have millisecond time resolution, they currently lack the ability to determine accurately where the recorded cells are. Ideally, the best approach is to perform statistical analysis using both electrophysiology and functional imaging in the same area; that way, both high temporal and high spatial resolution could be achieved.

It is important to note that the framework of interventional intersectional information requires knowing precisely which values of the neural response features r_i are elicited by intervention in each trial. This in turn requires measurement of the neural response (r_i) on individual intervention trials. When this is not possible, confounds may arise. For example, in the absence of such measures it would be problematic to rule out a residual correlation between the interventionally elicited neural response and other uncontrolled endogenous brain activity variables that would invalidate the rigor of the causal conclusions, or the elicited activity may be so un-natural (e.g., too synchronized with respect to natural activity patterns) that they may affect in an

un-natural way downstream neural processes. When it is not possible to measure in each trial the elicited neural response, the study should be however accompanied by a rigorous quantification in separate trials or experiments of the precision of manipulated response under various conditions, that is adequate to allow extrapolation to individual trials during the behavioral task and that also characterizes the difference between manipulated and non-manipulated activity.

Potential Confounds: When the Framework May Fail

The result of the intersection framework (and of any experimental approach combining neural recordings and interventional techniques) are potentially confounded by many limitations and factors that must be considered carefully to avoid reaching the wrong conclusions. We have already discussed some of those confounds; in the following we discuss additional ones.

A key requirement for the intersection framework to succeed is that the animal uses the identified stimulus to make choices. This requirement can fail in two important ways. First, in some behaviors, there may be other sensory stimuli that co-vary with the stimuli of interest. In this case, it would be difficult to know which stimulus feature is being used by the animal to drive behavior, a problem exacerbated by the possibility that the stimulus features used by the animal might vary from trial to trial. Second, the animal might have fluctuations in attention, motivation, or arousal, or use non-stimulus features such as reward history, to drive choices. These factors may be present in the two-alternative forced-choice tasks that we discussed in this article, but are likely to be stronger in other task designs, such as go/no go tasks, where our framework could be in principle applied. In all these cases, factors other than the stimulus feature of interest would be involved in driving the animal's choice; that would compromise the proposed framework, since it assumes that the stimulus of interest drives behavior. Such factors can be conceptually formalized by assuming that both the sensory coding and the decision mechanisms may vary across trials, and/or that non-recorded or non-manipulated neurons, may vary across intervention and non-intervention trials (such as r_2 depicted in Figures 4E and 4F when only intervening on feature r_1).

For these reasons, it is important to evaluate whether the variables describing behavior and the non-observed and non-manipulated endogenous variables are in a comparable state during intervention and non-intervention trials. In the presence of variations, a simple strategy could be to down-sample intervention and sensory-evoked trials so that only compatible brain or behavioral states are analyzed. A better solution, however, is to consider tasks in which it is known, based on high behavioral performance and good psychometric curves, that the stimulus feature of interest drives the animal's choice with high reliability. Similarly, the stimulus should be designed so that co-varying stimulus features are avoided. This will probably be easier with simple stimulus sets than with natural stimuli.

Variation of behavior and brain state variables across the experiment, on the other hand, offer an important opportunity to evaluate whether such variables have a "permissive" role on task performance. For example, in the virtual-pole sensation experiment of (O'Connor et al., 2013), the fact that virtual pole perception worked only when the animal whisked suggests a permissive role of whisker movements for active sensation. A strategy that could take advantage of these variations in state and behavior could be to include (using e.g., simple modeling techniques such as Generalized Linear Models [Park et al., 2014]) behavioral factors such as slow variation across blocks of trials of motivation or reward history or brain states explicitly into the experimenter's sensory coding and decision boundary models. This could potentially lead to explaining the dynamic role of these factors in sensory coding.

Another significant confound can arise for interventional approaches when investigating partly parallel pathways. For example, suppose a behavior is generated by two brain areas that operate in a partly parallel or complementary way, as for example in an "OR" function (Li et al., 2016). In such a case, when inactivating only one area with intervention, one may find little causal effect on behavior. However, interpreting this result as evidence that the inactivated region does not causally contribute much to behavior could be misleading. One way to alleviate these confounds would be to compute both statistical and interventional information intersection. One may use these measures to disambiguate the case in which the two areas contribute complementarily to behavior and so offer complementary intersection information from the case when the two areas operate entirely redundantly and so intersection information from both areas equals intersection information from one area alone. However, accurate interpretation of these measures would require knowledge of the functional anatomy, which, for example, informs the experimenter about the presence or absence of parallel and potentially redundant pathways. Moreover, completeness of activity monitoring and perturbation of the regions involved is also paramount, as this would be, for instance, useful to rule out that a failure to affect behavior by inactivating a region is due to incomplete control of all relevant neurons. We thus anticipate that as approaches move toward understanding larger and larger populations of neurons (Keller and Ahrens, 2015; Sofroniew et al., 2016) and the interconnections between neurons (Lichtman and Denk, 2011), these joint statistical and interventional approaches will become easier to interpret.

Determining the Instructive versus Permissive Role of Neural Codes and Neural Circuit: From Circuit Dissection to Circuit Information Flow

The intersection information framework (both statistical and interventional) has direct application for the dissection of neural circuits underlying behavior. Much work in systems neuroscience has used neurophysiology to identify neural correlates, and, due to recent optogenetics approaches, a wave of new studies has sought to identify which brain regions, cell types, axonal projection pathways, and circuits are required for accurate performance of behavioral tasks (Guo et al., 2014; O'Connor et al., 2013; Peng et al., 2015). It is essential to emphasize that simply measuring the effect of an intervention on choice without regard to stimulus coding precludes understanding a neural circuit's role in task performance. Here we propose that the use of intersection information is crucial to determine whether a neural circuit (or cell type or projection pathway) carries information that is instructive (Otchy et al., 2015) for task performance

(contributes essential information for the task performance that is not provided elsewhere) or if the circuit is permissive for task performance (is required for, or modulates, the behavior but does not provide essential information).

Figures 4A–4D schematizes four cases of different neural circuit architectures in which two neural features, r_1 and r_2 (for example, the activity of neurons in two different brain areas, cell types, or projection pathways), may inform choice. In all four cases, feature r_2 contributes essential information to choice and task performance (r_2 is thus instructive), but the role of r_1 varies. An interventional intersection framework would correctly identify these four circuit architectures. The case of "parallel" information flow (Figure 4A), in which r_1 and r_2 both provide complementary instructive information, could be revealed by finding that the intersection information provided by r_1 and r_2 jointly is larger than that provided by r_1 alone (that is, r_1 and r_2 provide complementary stimulus information to choice). The case of "serial" information flow (Figure 4C), in which r_1 provides instructive information to r_2 and r_2 informs choice, could be discovered by finding that the intersection information provided by r_1 and r_2 jointly equals that provided by r_2 alone. The case in which r_1 provides permissive—but not instructive—information (Figure 4D), could be identified by finding that r_1 carries interventional choice information but not intersection information. Finally, the case when r_1 is not used for choice (Figure 4B) corresponds to the absence of both choice and intersection information in r_1.

It is important to note that the framework we discussed here is general and can in principle be applied not only to determine how sensory information carried by different codes is used to produce behavior, but it can also be used to study how stimulus information flows across neural populations. For example, the same reasoning expressed above applies to considering a group of brain regions $r_1,\ldots,r_n$ (whose activity we can record and manipulate, not necessarily at the same time) and a downstream area c (whose activity we assume we can record at the same time when we manipulate or record $r_1,\ldots,r_n$). In this case, the meaning of intersection information would be that of the amount of information about stimulus s carried by population $r_1,\ldots,r_n$ that is transmitted downstream to area c. In essence, we have replaced choice by activity in area c. We could therefore identify the neural response features that influence activity in downstream regions, leading to hypotheses about the mechanisms of information flow in neural circuits.

Ideally, these statistical and causal measures of information flow should be integrated with information about anatomy, response timing, and information dynamics. For example, in the presence of a partly feedforward or hierarchical architecture, anatomy could be used to identify the earliest areas where sensory, choice, and intersection information are developed (Koulakov et al., 2005), and thus better track the computations leading to task performance. Similarly, the timing of stimulus and choice information in neural activity could be used to infer whether, for example, choice signals reflect a neuron's causal effect on behavioral choice or rather a top-down signal (Nienborg and Cumming, 2009).

Concluding Remarks

We presented a new framework to crack the neural code underlying sensory perception. The framework emphasizes neural response features that both carry sensory information and lead to appropriate actions, with the emphasis on "appropriate." These are the neural response features with a large intersection between sensory information and readout. Based on this framework, we provided an initial attempt to formalize statistical ways to identify these features from recordings of neural activity, and to design interventional experiments that can causally test the degree of intersection information. This approach can resolve open debates about the nature of the neural code. Moreover, the ideas we proposed in this framework can guide researchers in the design of experiments, in the design of new statistical tools, and in the development of the new technology, that will lead us to crack the neural code.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and one figure and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2016.12.036.

ACKNOWLEDGMENTS

We thank the referees for their insightful comments, and members of our laboratories for precious feedback. We are particularly indebted to M.E. Diamond, H. Safaai, D. Chicharro, C. Kayser, and C. Runyan for valuable collaborations on these topics and/or comments on the manuscript. S.P. and E.P. acknowledge the warm hospitality of Harvard Medical School while writing this paper. This work was supported by the Fondation Bertarelli, the European Research Council (ERC, NEURO-PATTERNS), the Flag-Era JTC Human Brain Project (SLOW-DYN), a Burroughs-Wellcome Fund Career Award at the Scientific Interface, the Searle Scholars Program, the New York Stem Cell Foundation, and NIH grants R01-MH107620, R01-NS089521, and U01NS090576. C.D.H. is a New York Stem Cell Foundation Robertson Neuroscience Investigator. Funding for P.E.L. was provided by the Gatsby Charitable Foundation.

REFERENCES

Ahmadian, Y., Packer, A.M., Yuste, R., and Paninski, L. (2011). Designing optimal stimuli to control neuronal spike timing. J. Neurophysiol. *106*, 1038–1053.

Averbeck, B.B., Latham, P.E., and Pouget, A. (2006). Neural correlations, population coding and computation. Nat. Rev. Neurosci. *7*, 358–366.

Baden, T., Berens, P., Franke, K., Román Rosón, M., Bethge, M., and Euler, T. (2016). The functional diversity of retinal ganglion cells in the mouse. Nature *529*, 345–350.

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*, 5.

Bovetti, S., and Fellin, T. (2015). Optical dissection of brain circuits with patterned illumination through the phase modulation of light. J. Neurosci. Methods *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. Nat. Neurosci. *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.

Buonomano, D.V., and Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. Nat. Rev. Neurosci. *10*, 113–125.

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

Chen, X., Gabitto, M., Peng, Y., Ryba, N.J.P., and Zuker, C.S. (2011). A gustotopic map of taste qualities in the mammalian brain. Science *333*, 1262–1266.

Chen, J.L., Carta, S., Soldado-Magraner, J., Schneider, B.L., and Helmchen, F. (2013). Behaviour-dependent recruitment of long-range projection neurons in somatosensory cortex. Nature *499*, 336–340.

Cunningham, J.P., and Yu, B.M. (2014). Dimensionality reduction for large-scale neural recordings. Nat. Neurosci. *17*, 1500–1509.

Deisseroth, K., and Schnitzer, M.J. (2013). Engineering approaches to illuminating brain structure and dynamics. Neuron *80*, 568–577.

Doron, G., von Heimendahl, M., Schlattmann, P., Houweling, A.R., and Brecht, M. (2014). Spiking irregularity and frequency modulate the behavioral report of single-neuron stimulation. Neuron *81*, 653–663.

Emiliani, V., Cohen, A.E., Deisseroth, K., and Häusser, M. (2015). All-Optical Interrogation of Neural Circuits. J. Neurosci. *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. Nat. Neurosci. *11*, 603–608.

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

Guo, Z.V., Li, N., Huber, D., Ophir, E., Gutnisky, D., Ting, J.T., Feng, G., and Svoboda, K. (2014). Flow of cortical activity underlying a tactile decision in mice. Neuron *81*, 179–194.

Haefner, R.M., Gerwinn, S., Macke, J.H., and Bethge, M. (2013). Inferring decoding strategies from choice probabilities in the presence of correlated variability. Nat. Neurosci. *16*, 235–242.

Harvey, C.D., Coen, P., and Tank, D.W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature *484*, 62–68.

Harvey, M.A., Saal, H.P., Dammann, J.F., 3rd, and Bensmaia, S.J. (2013). Multiplexing stimulus information through rate and temporal codes in primate somatosensory cortex. PLoS Biol. *11*, e1001558.

Hernández, A., Nácher, V., Luna, R., Zainos, A., Lemus, L., Alvarez, M., Vázquez, Y., Camarillo, L., and Romo, R. (2010). Decoding a perceptual decision process across cortex. Neuron *66*, 300–314.

Histed, M.H., and Maunsell, J.H. (2014). Cortical neural populations can guide behavior by integrating inputs linearly, independent of synchrony. Proc. Natl. Acad. Sci. USA *111*, E178–E187.

Houweling, A.R., and Brecht, M. (2008). Behavioural report of single neuron stimulation in somatosensory cortex. Nature *451*, 65–68.

Huber, D., Petreanu, L., Ghitani, N., Ranade, S., Hromádka, T., Mainen, Z., and Svoboda, K. (2008). Sparse optical microstimulation in barrel cortex drives learned behaviour in freely moving mice. Nature *451*, 61–64.

Jacobs, A.L., Fridman, G., Douglas, R.M., Alam, N.M., Latham, P.E., Prusky, G.T., and Nirenberg, S. (2009). Ruling out and ruling in neural codes. Proc. Natl. Acad. Sci. USA *106*, 5936–5941.

Jia, X., Tanabe, S., and Kohn, A. (2013). γ and the coordination of spiking activity in early visual cortex. Neuron *77*, 762–774.

Kayser, C., Montemurro, M.A., Logothetis, N.K., and Panzeri, S. (2009). Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. Neuron *61*, 597–608.

Keller, P.J., and Ahrens, M.B. (2015). Visualizing whole-brain activity and development at the single-cell level using light-sheet microscopy. Neuron *85*, 462–483.

Koulakov, A.A., Rinberg, D.A., and Tsigankov, D.N. (2005). How to find decision makers in neural networks. Biol. Cybern. *93*, 447–462.

Lerner, T.N., Ye, L., and Deisseroth, K. (2016). Communication in Neural Circuits: Tools, Opportunities, and Challenges. Cell *164*, 1136–1150.

Li, N., Chen, T.W., Guo, Z.V., Gerfen, C.R., and Svoboda, K. (2015). A motor cortex circuit for motor planning and movement. Nature *519*, 51–56.

Li, N., Daie, K., Svoboda, K., and Druckmann, S. (2016). Robust neuronal dynamics in premotor cortex during motor planning. Nature *532*, 459–464.

Lichtman, J.W., and Denk, W. (2011). The big and the small: challenges of imaging the brain's circuits. Science *334*, 618–623.

Luczak, A., McNaughton, B.L., and Harris, K.D. (2015). Packet-based communication in the cortex. Nat. Rev. Neurosci. *16*, 745–755.

Luna, R., Hernández, A., Brody, C.D., and Romo, R. (2005). Neural codes for perceptual discrimination in primary somatosensory cortex. Nat. Neurosci. *8*, 1210–1219.

Meister, M.L., Hennig, J.A., and Huk, A.C. (2013). Signal multiplexing and single-neuron computations in lateral intraparietal area during decision-making. J. Neurosci. *33*, 2254–2267.

Morcos, A.S., and Harvey, C.D. (2016). History-dependent variability in population dynamics during evidence accumulation in cortex. Nat. Neurosci. *19*, 1672–1681.

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

Nienborg, H., and Cumming, B.G. (2009). Decision-related activity in sensory neurons reflects more than a neuron's causal effect. Nature *459*, 89–92.

Nikolić, D., Fries, P., and Singer, W. (2013). Gamma oscillations: precise temporal coordination without a metronome. Trends Cogn. Sci. *17*, 54–55.

O'Connor, D.H., Hires, S.A., Guo, Z.V., Li, N., Yu, J., Sun, Q.Q., Huber, D., and Svoboda, K. (2013). Neural coding during active somatosensation revealed using illusory touch. Nat. Neurosci. *16*, 958–965.

O'Keefe, J., and Recce, M.L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. Hippocampus *3*, 317–330.

Otchy, T.M., Wolff, S.B.E., Rhee, J.Y., Pehlevan, C., Kawai, R., Kempf, A., Gobes, S.M.H., and Ölveczky, B.P. (2015). Acute off-target effects of neural circuit manipulations. Nature *528*, 358–363.

Packer, A.M., Russell, L.E., Dalgleish, H.W.P., and Häusser, M. (2015). Simultaneous all-optical manipulation and recording of neural circuit activity with cellular resolution in vivo. Nat. Methods *12*, 140–146.

Panzeri, S., Brunel, N., Logothetis, N.K., and Kayser, C. (2010). Sensory neural codes using multiplexed temporal scales. Trends Neurosci. *33*, 111–120.

Papagiakoumou, E., Anselmi, F., Bègue, A., de Sars, V., Glückstad, J., Isacoff, E.Y., and Emiliani, V. (2010). Scanless two-photon excitation of channelrhodopsin-2. Nat. Methods *7*, 848–854.

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

Pearl, J. (2009). Causality (Cambdrige University Press).

Peng, Y., Gillis-Smith, S., Jin, H., Tränkner, D., Ryba, N.J.P., and Zuker, C.S. (2015). Sweet and bitter taste in the brain of awake behaving animals. Nature *527*, 512–515.

Peron, S., and Svoboda, K. (2011). From cudgel to scalpel: toward precise neural control with optogenetics. Nat. Methods *8*, 30–34.

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

Reich, D.S., Mechler, F., and Victor, J.D. (2001). Independent and redundant information in nearby cortical neurons. Science *294*, 2566–2568.

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

Rigotti, M., Barak, O., Warden, M.R., Wang, X.J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature *497*, 585–590.

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

Romo, R., Hernández, A., Zainos, A., and Salinas, E. (1998). Somatosensory discrimination based on cortical microstimulation. Nature *392*, 387–390.

Salzman, C.D., Britten, K.H., and Newsome, W.T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. Nature *346*, 174–177.

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. J. Neurosci. *16*, 1486–1510.

Shamir, M. (2014). Emerging principles of population coding: in search for the neural code. Curr. Opin. Neurobiol. *25*, 140–148.

Shriki, O., Kohn, A., and Shamir, M. (2012). Fast coding of orientation in primary visual cortex. PLoS Comput. Biol. *8*, e1002536.

Shusterman, R., Smear, M.C., Koulakov, A.A., and Rinberg, D. (2011). Precise olfactory responses tile the sniff cycle. Nat. Neurosci. *14*, 1039–1044.

Sofroniew, N.J., Flickinger, D., King, J., and Svoboda, K. (2016). A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. eLife *5*, 5.

Sternson, S.M., and Roth, B.L. (2014). Chemogenetic tools to interrogate brain functions. Annu. Rev. Neurosci. *37*, 387–407.

Tehovnik, E.J., Tolias, A.S., Sultan, F., Slocum, W.M., and Logothetis, N.K. (2006). Direct and indirect activation of cortical neurons by electrical microstimulation. J. Neurophysiol. *96*, 512–521.

Victor, J.D. (2000). How the brain uses time to represent and process visual information(1). Brain Res. *886*, 33–46.

Victor, J.D., and Nirenberg, S. (2008). Indices for testing neural codes. Neural Comput. *20*, 2895–2936.

Wilson, N.R., Runyan, C.A., Wang, F.L., and Sur, M. (2012). Division and subtraction by distinct cortical inhibitory networks in vivo. Nature *488*, 343–348.

Woods, A.J., Antal, A., Bikson, M., Boggio, P.S., Brunoni, A.R., Celnik, P., Cohen, L.G., Fregni, F., Herrmann, C.S., Kappenman, E.S., et al. (2016). A technical guide to tDCS, and related non-invasive brain stimulation tools. Clin. Neurophysiol. *127*, 1031–1048.

Yang, Y., and Zador, A.M. (2012). Differences in sensitivity to neural timing among cortical areas. J. Neurosci. *32*, 15142–15147.

Yang, Y., DeWeese, M.R., Otazu, G.H., and Zador, A.M. (2008). Millisecond-scale differences in neural activity in auditory cortex can drive decisions. Nat. Neurosci. *11*, 1262–1263.

Zuo, Y., 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. Curr. Biol. *25*, 357–363.

Neuron, Volume *93*

Supplemental Information

Cracking the Neural Code for Sensory Perception by Combining Statistics, Intervention, and Behavior

Stefano Panzeri, Christopher D. Harvey, Eugenio Piasini, Peter E. Latham, and Tommaso Fellin

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 $\boldsymbol{r} = (r_1, ..., r_n)$. 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

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

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 $\boldsymbol{r}$ 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:

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

A value of *II* higher than chance means 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

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

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 behaviorally 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

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

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:

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

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** by itself does not affect choice, but inherits choice selectivity by being correlated with a variable that affects choice (Ince et al., 2012). One possibility is the case plotted in Figs. 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 does affect 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 not depending on correctness of a 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 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 means the stimulus is invisible from the background and 100% signal intensity means 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 spike count (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

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

where

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

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

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

$$p(r_3 \mid \rho) = N\left(\rho, \sigma_3^2\right) \quad .$$

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

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

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 $(r_- + r_3) \sim N(\rho, \sigma_-^2 + \sigma_3^2)$, 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):

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

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). Moreover, 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

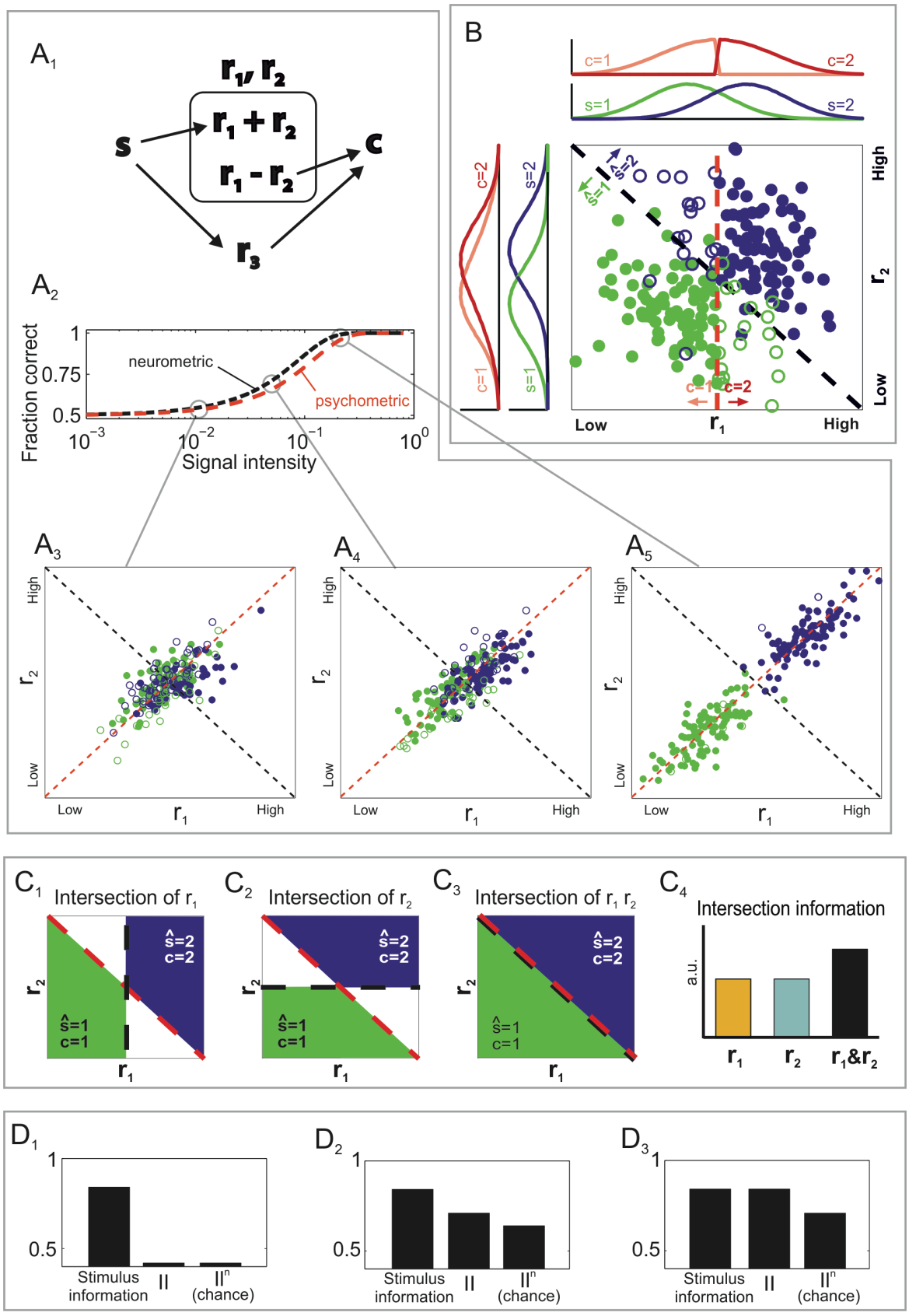
$$\boldsymbol{\Sigma} = \begin{pmatrix} 0.2 & -0.005 \\ -0.005 & 0.2 \end{pmatrix}$$

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 (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 USA 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, e14193

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 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 USA 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, 140-146.

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, 848-854.

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 USA 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 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.