Review:
Characterizing the influence of 'internal states' on sensory activity

Richard D. Lange* & Ralf M. Haefner*
Brain & Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA

March 6, 2017

Abstract

The concept of a tuning curve has been central for our understanding of how the responses of cortical neurons depend on external stimuli. Here, we describe how the influence of unobserved internal variables on sensory responses, in particular their response correlations, can be understood in a similar framework. We review related recent work and discuss its implication for the information contained in sensory responses.

Introduction

A large part of cortical function can be characterized as transforming sensory inputs to behavioral outputs. While this transformation is conceptually unidirectional, the anatomical structures implementing it are largely bidirectional [1]. In this review, we will concentrate on discussing progress in understanding feedback signals (FB) that influence neural responses in early visual areas, and their relation to classically feedforward (FF) models of sensory processing. We expect the discussed techniques and computational frameworks to also be useful in understanding neural responses in other sensory or motor areas.

The transformation of sensory inputs into behavioral outputs can be understood on at least two levels. On the first one, we would like to understand how a stimulus, S, influences neural responses, $\mathbf{r}$, and how these responses influence behavior, B. On a more abstract level, one can try to understand how a stimulus affects internal states, I, and how these internal states are linked to behavior (Figure 1). While those two levels are clearly linked, the representation of abstract internal states is generally unclear, and it is in principle only possible to directly observe the corresponding neural responses. Internal states are only accessible by their covariability with one or more of the observable quantities, S, B, and $\mathbf{r}$. In general, all of these quantities vary over time [2], and observations of their joint or conditional probability distributions can tell us about different aspects of brain function (Figure 1).[1]

$p(S, \mathbf{r})$: Characterizes the stimulus-dependence of neural responses, giving rise to tuning functions and receptive fields.

*{rlange,rhaefne2}@ur.rochester.edu

[1] We use joint distributions here because they allow us to characterize relationships even when one of the variables is not directly observed.

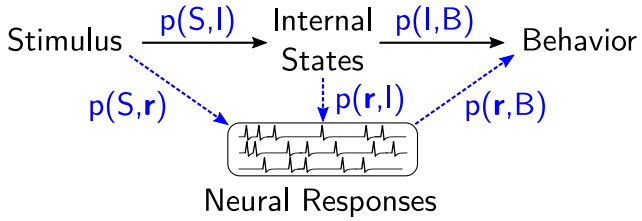

Figure 1: **a:** A computational-level description of the brain may invoke abstract *Internal States* (I) that govern behavior (B) and are influenced by stimuli (S). On a neurophysiological level, we also seek a description on the level of *responses* ($\mathbf{r}$) of populations of neurons. It is often useful to mix levels of abstraction, for example modeling the effect of attention (an abstract state) on neural responses.

$p(\mathbf{r}, B)$: Usually called 'decision-related' or 'choice-related' activity in the context of decision-making tasks [3]. When $\mathbf{r}$ is measured in motor cortex, it gives rise to tuning curves with respect to motor outputs [4, 5].

$p(\mathbf{r}, I)$: Depending on the nature of I, this can alternatively be interpreted as the influence of cortical states on neural responses (e.g. anesthesia, attentional state, motor activity) [6, 7, 8, 9$^{\bullet\bullet}$, 10, 11$^{\bullet\bullet}$, 12, 13], or their neural representation (e.g. for beliefs about the outside world) [14, 15, 16$^{\bullet}$, 17$^{\bullet}$].

$p(S, I)$: Not observable directly, but may be induced by experimental design, for instance by training a subject to allocate attention following an external cue, or when measuring neural correlates of reward and value.

$p(I, B)$: Captures differences in behavior depending on internal state, e.g. performance differences between attentional states, or different choices depending on percept. Again, not directly observable, but usually implicitly assumed to be under the control of the experimenter.

We will first describe the implications of extending the concept of a tuning function to characterizing the influence of internal states on neural responses. We will then describe recent efforts to distinguish between feedforward and feedback influences on neural responses, and the implications for the information represented by sensory neurons.

1 Characterization of the influence of internal states

Box 1: Tuning and Covariability

Just as the dependence of a neuron's firing rate on an *external* parameter yields a tuning curve, a similar dependence on some *internal* state may be thought of as "tuning" to that internal state, though the value of that state may not be precisely known [2]. Taken together, we can write a population's response as a vector-valued 'tuning' function of the external parameters and internal states on which it depends[†]:

$$\mathbf{r} = \mathbf{f}(\underbrace{a, b, \ldots}_{\text{external}}, \underbrace{x, y, \ldots}_{\text{internal}}) + \text{noise} \tag{1}$$

Taking the linear approximation of $\mathbf{f}$ (Figure 2), the covariance between two neurons' responses takes the form of a sum of outer products [18•, 19••, 16•]:

$$\mathbf{C} \approx \mathbf{C}^0 + \underbrace{\sum_{v \in \text{vars}} \frac{\mathrm{d}\mathbf{f}}{\mathrm{d}v}\frac{\mathrm{d}\mathbf{f}}{\mathrm{d}v}^{\top} \mathrm{var}(v)}_{\text{independent terms}} + \underbrace{\sum_{u,v \in \text{vars}} \left(\frac{\mathrm{d}\mathbf{f}}{\mathrm{d}u}\frac{\mathrm{d}\mathbf{f}}{\mathrm{d}v}^{\top} + \frac{\mathrm{d}\mathbf{f}}{\mathrm{d}v}\frac{\mathrm{d}\mathbf{f}}{\mathrm{d}u}^{\top} \right) \mathrm{cov}(u, v)}_{\text{interaction terms}} \tag{2}$$

where $\mathbf{C}^0$ is 'intrinsic' covariance (i.e. the covariance of the noise) and $u, v \in$ vars represents any of the internal or external variables that f depends on.

Equation (2) makes explicit how shared sources of variability introduce structured co-variability into a population. Importantly, both variation in *internal* states and variation in *external* stimulus parameters affect a population's covariance analogously. By analyzing a population's covariance structure, then, we can learn about both stimuli and internal states modulating neural responses.

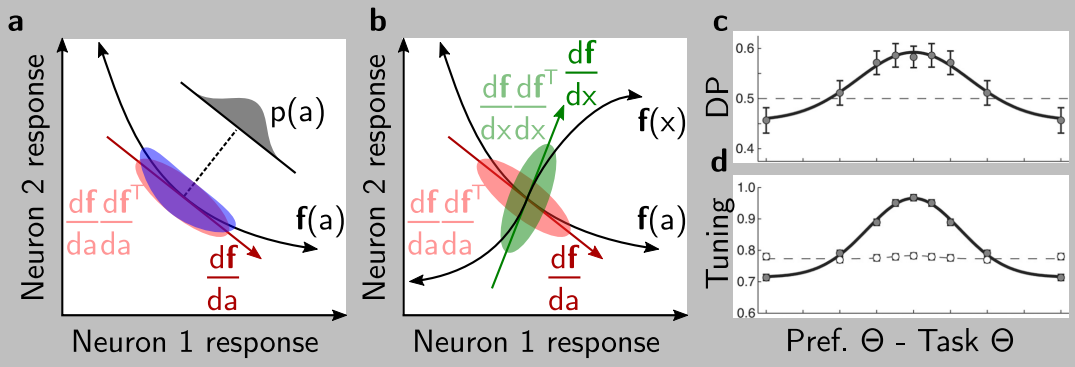

Figure 2: **a:** The population tuning of two neurons to some parameter, $\mathbf{f}(a)$, entails response co-variability along $\mathbf{f}$ (blue) [18•], often called 'signal correlations' when a reflect the stimulus, and 'noise correlations' when it reflects uncontrolled internal variability [20, 21]. **b:** Variability in multiple parameters adds (eq. 2). **c-d:** Population responses of [22] replotted to illustrate conceptual similarity between tuning with respect to outside stimulus (d), and tuning with respect to behavior (c). Here, monkeys performed a motion-based detection task while the authors recorded from macaques' MT neurons. **c:** Dependence of a neuron's 'detect probability' (DP) on its preferred motion direction relative to the task direction measured. DP is closely related to the difference in the mean responses associated with the two choices, $\Delta\mathbf{f}/\Delta$choice [23]. Hence, if DPs reflect FB influences from a decision-making area[24••, 25], then this curve is interpretable as *tuning to the decision state*. **d:** Normalized difference in response associated with the two task-relevant stimuli, $\Delta\mathbf{f}/\Delta$stim, showing the relationship between tuning to internal and to external parameters.

[†]Correlations between I and $\mathbf{r}$ considered here may reflect causation from I to $\mathbf{r}$, or common input to both. Equation (2) holds as long as I is not purely a function of $\mathbf{r}$, in which case variability in I will not affect variability in $\mathbf{r}$.

Stimulus-dependence of responses $p(S, \mathbf{r})$: Observing the full joint distribution $p(S, \mathbf{r})$ is rarely possible even for a single neuron due to the high dimensionality of S, at least in the visual domain. As a result, experimentalists have often concentrated on probing this relation along a single dimension while keeping everything else about the stimulus constant. For example, varying the orientation of a grating pattern yields the orientation tuning curve of each recorded neuron. In general, a neuron may be tuned to multiple independent parameters. For example, V1 neurons are often characterized jointly by their tuning to the orientation and to the spatial-frequency of a grating stimulus.

Behavior-related responses $p(\mathbf{r}, B)$: Taking the same approach as above, we can use concurrent observations of neural responses and behavioral outputs to infer their relationship. Traditionally, this has been done by computing 'choice probability' and 'detect probability' [26, 25, 27] – measures for how neural responses differ depending on behavioral response. A closely related but simpler measure is the dependence of the *mean* neural response on the behavioral response. For continuous behavioral outputs such as eye or limb movements [28, 29], such a measure is equivalent to the classic concept of a tuning curve. For binary responses as common in perceptual decision-making tasks (reviewed in [25]), it is closely related to choice and detect probabilities under reasonable assumptions [30].

In a purely feedforward, unidirectional model, the dependence of neural responses on behavior would reflect the decoding strategy of the brain, i.e. how the neural responses inform behavior. However, in a bidirectional model including internal variables that influence both responses and outputs, this dependence is best thought of as a covariance between responses and outputs, and does not imply only feedforward causation (Box 1).

Internal state-related responses $p(\mathbf{r}, I)$: Since a subject's internal state is not observable directly, the traditional approach to measuring its impact on neural responses has been to exploit its relationship to observed quantities like stimulus and behavior. For example, training an animal in an attention task induces a relationship between an aspect of the internal state ('attentional state') and an external sensory signal ('attention cue') [2], where improved behavioral performance is taken as an indicator for successful manipulation of attentional state [9••]

There are two limitations to this approach. (1) there usually remains significant uncertainty about the relationship between observables and internal states. For instance, even though an attention cue may only have two states, the internal variables determining the animal's attentional state is likely non-binary and more variable. And (2), the internal state is much higher-dimensional than the observations. For instance, the attentional state is likely high-dimensional determining location(s), feature(s), etc. that the animal is attending to [9••, 31, 19••] (Box 2). Equally, the internal percept is surely richer than the binary output in a perceptual decision-making task. Both problems severely limit the information that can be inferred about internal states and their influence on neural responses from their dependence on the preceding stimulus or their dependence on the behavioral report.

What covariability reveals about internal states: The above approaches can all be considered *conditional*: they rely on measuring each neuron's response conditioned on other quantities, e.g. conditioned on choice or on the attentional cue. One way to overcome the described limitations is by considering the *full* response distribution, i.e. across all choices or cues. If some component, x, of the subject's internal state varies uncontrolled from trial to trial, the population response $\mathbf{r}$ moves together according to its 'tuning' to x (Box 1). This means that variability in an internal

state leaves a signature in the statistical structure of the neural population response. Furthermore, different internal states all leave their own signatures that will superpose (Box 2). By recording from sufficiently many neurons concurrently, and by using sufficiently powerful statistical inference techniques, both of which are under current development [32•, 33, 34], we can infer *both* the internal states influencing neural responses, and how those responses depend on them [9••, 31, 35••]. This will also allow estimation of the internal state as a function of time [11••] and on individual trials [9••]. Furthermore, with nonlinear dimensionality reduction methods one can infer curvature in $\mathbf{f}$ as well, going beyond the simple but instructive picture of linear tuning presented above [32•].

Experimental data: A number of papers can be interpreted in this framework as having measured consequences of tuning curves with respect to internal variables by correlating neural responses with behavior [22, 36, 37, 38••, 9••, 31, 11••, 35••, 39••], four of which we will highlight here (also see Fig. 2 in Box 1, and Box 2).

Having trained a monkey on a cued change-detection task, [9••] computed the line connecting the mean neural response in attention trials with the mean neural response in the attend-away trials. This is equivalent to computing $\mathbf{f}'$ with respect to attention. The authors overcome the challenge that attentional state is not precisely known by binarizing it ('attend left' or 'attend right') and discarding false negative trials, where attention is assumed to not match the cued location. They found that their trial-by-trial estimates of attentional state predicted performance on individual trials, and that it varied substantially across trials even within the same experimental condition.

[11••] regressed a scalar latent factor influencing neural activity in primate V1 both during anesthesia and while awake. The shared, time-varying gain term accounted for a large part of the population's variability as well as their correlations during the anesthetized, but not during the awake state. This is an example of inferring an internal state which is clearly important but for which no normative account exists at present.

[35••] explicitly modeled the trial-by-trial changes in V4 spiking activity during a change-detection task as a combination of a stimulus-driven component and shared time-varying feedback terms that were either known (the attentional cue) or fit to the data. The inferred values of the fitted internal states were then correlated per-trial with spiking statistics ($p(I, \mathbf{r})$) and the animal's behavior ($p(I, B)$), and were found to explain a large amount of variance in both.

At least two groups have measured the effect of changing the task context on noise correlations while keeping the stimulus distribution the same [38••, 39••]. [38••] find, in a change detection task, correlations that are well-described as the effect of only two varying attentional states (Box 2). [39••] measured the noise correlation matrix in V1 while the monkey performed a coarse discrimination task. Interestingly, they found that the feedback component of correlations could largely be explained by variability in only a single state. Furthermore, [39••] were able to isolate and quantify the task-dependent component of both choice probabilities and noise correlations, finding a significant contribution of task-related covariability to both.

Box 2: Case Study: [38••]

In Box 1 we argued that response correlations (or more precisely, covariances) may be usefully understood as the sum of 'intrinsic' covariability with outer products of the population's 'tuning' to each independently varying parameter. [38••] used essentially this idea, which we replicate and make more explicit here. The authors measured pairwise noise correlations between MT neurons in a motion discrimination task, where the discriminated directions changed from trial to trial. Recording from pairs of neuron at the same time, switching tasks implied that firing by the two neurons either supported the same decision, or opposite decisions. Their results showed noise correlations change depending on whether the neurons support the same choice or different choices, implying an influence of the internal state 'task context' on MT responses (Figure 3, bottom left).

Following [38••], we assume two independently varying internal states, x and y: one for alternating 'attention' between the discriminated directions, and one for fluctuating 'attention' to both of the task-relevant motion directions concurrently [19••].

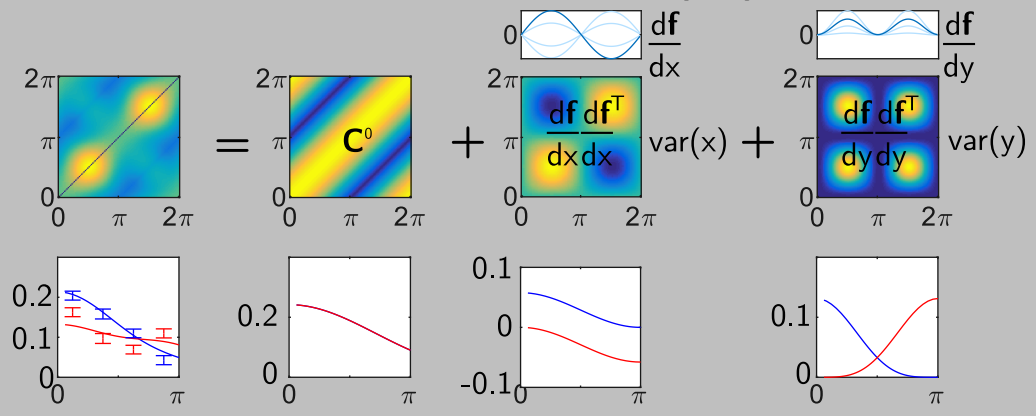

Figure 3: **Middle row:** a full simulated correlation matrix, plotted as a function of each neuron's preferred motion direction, is derived from the weighted sum of independent variability (not shown), limited-range covariance, and independent outer-product terms (Equation 2). **Bottom row:** [38••] measured pairwise correlations as a function of the difference in the pair's preferred directions, ranging from 0 to 180°. Data replotted as error bars and compared to model (lines). Pairs where neurons support the same choice are in blue, and pairs supporting different choices are in red. Both, the initial separation between the conditions as well as the crossover at larger differences in preferred direction are captured by the above two components of attention.

2 Distinguishing feedforward from feedback influences

Because the external stimulus is under the direct control of the experimenter, it is possible to establish a causal feedforward relationship between external inputs and the responses of sensory neurons. In order to establish causality in the relationships between neural responses and behavior, or neural responses and internal states, it is likewise necessary to experimentally manipulate either of them. Two non-causal approaches to distinguish between feedforward and feedback signals have been (1) to compare the observations to the predictions from a purely feedforward model and ascribe any residuals to feedback influences, and (2) to compare neural responses for different internal states under the *assumption* that those states are represented by neurons downstream of those that are

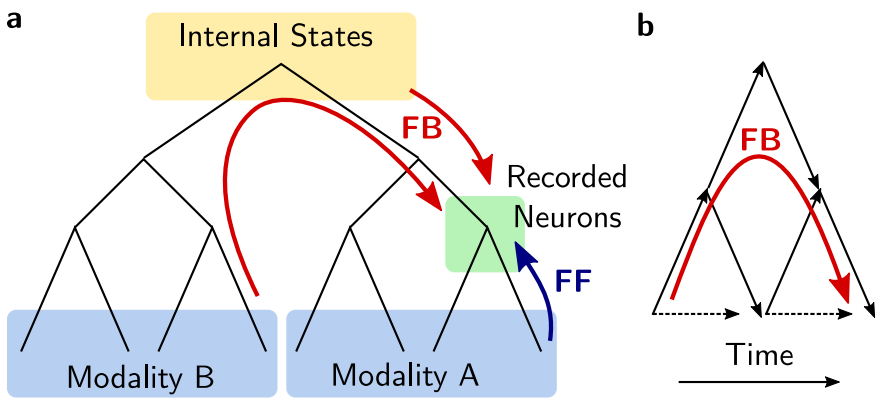

Figure 4: Identification of feedback signals by dependency of neural response on stimuli not within the presumed feedforward pathway. **a:** Dependency of neural responses in other sensory modalities, or outside the receptive field. **b:** Dependency on external inputs at different times, sufficiently long in the past such that 'memory' about them within the feedforward pathway can be excluded.

being observed (Figure 4). Note that we refer here only to the functional definition of FB, and that FB signals from downstream neurons may themselves not involve just pure cortico-cortico FB pathways but also anatomically FF pathways [40], for instance via a thalamocortical loop [41, 42].

[24••] took the first of these approaches. In monkeys performing a disparity discimination task, they compared the relationships between stimulus and choice to the relationship between sensory area V2 neural responses and choice. They found that the stimulus was most strongly correlated with the choice at the beginning of a trial, while the neural responses were most strongly correlated with the choice towards the end of the trial. Assuming that correlations between the neural responses do not change over the course of a trial one would expect the time courses of stimulus–choice, and response–choice correlations to agree, contrary to what is observed [43]. The authors therefore concluded that the neural responses they observed must be partly due to non-feedforward influences, most notably feedback from downstream decision-related neurons. Since it is conceivable that either feedforward inputs or recurrent connectivity among the sensory neurons could lead to increasing response correlations, answering this question definitively requires direct observations of those correlations.

Most studies taking the second approach manipulate the attentional state of the subject, e.g. by presenting a pre-trial cue, and have been extensively reviewed elsewhere [44]. An alternative approach is to analyze the co-variability of the neural responses and determine whether they contain task-dependent components. Under the critical assumption that task-learning, or task-switching, does not alter the feedforward connectivity, or recurrent connectivity, one can then conclude that any such task-dependent component must be due to feedback signals. The underlying assumption is more likely to be true for early sensory areas than higher levels of sensory processing where learning-induced plasticity has observed to be stronger [45]. The assumption is also more likely to be true for task-switches across shorter time scales, e.g. from trial to trial (seconds) rather than days or weeks over which relevant changes of the feedforward pathway become more likely. Both [38••] (see Box 2) and Bondy & Cumming [39••] have taken this approach comparing changes across trials in MT responses, and changes separated by several days in V1 responses, respectively.

Even though [35••] did not explicitly vary the task, a similar logic can be applied: Since the main common factors driving trial-to-trial variability impact individual neurons depending on their *task-dependent* tuning properties, they have to be different if the task is changed, and hence they

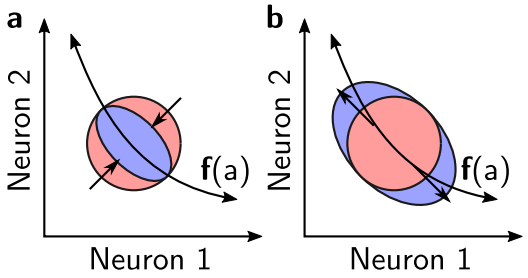

Figure 5: $d'd'-$correlations ($d' \equiv f'/\sigma$) can arise in different ways, with differing implications for sensory information which is determined by covariance. The blue response distributions have the same correlation but different covariances. **a:** No change in sensory information if $d'd'-$correlation is induced by suppression of variability orthogonal to $\mathbf{f}'$. This case implies lower response variability for both neurons. **b:** Decrease in sensory information if $d'd'-$correlations is induced by increased variability along $\mathbf{f}'$, e.g. by variability in an internal state whose tuning is aligned with $\mathbf{f}'$, *assuming that this internal state contains no information about* a.

must be due to feedback signals.

In this framework, training on a task induces a task-specific association or relationship between internal states like choice and sensory responses. If those internal states are represented by downstream neurons, and if this relationship is purely feedforward in nature, then neural responses will affect those states, and in turn observable behavioral outputs, but not the other way around. If, however, those internal states affect sensory responses, then sensory responses will differ depending on the task.

3 Implication for the information in sensory responses

Fisher information is often used to quantify the amount of information contained in sensory population responses. It is important to note that this information is computed not with respect to the entire input signal, S, but with respect to an experimenter-defined aspect, a, of it (e.g. with respect to orientation instead of the entire retinal image). In general, response variability that is indistinguishable from the variability along $\frac{\mathrm{d}\mathbf{f}}{\mathrm{d}a}$ that would be induced by variability in the relevant aspect of the external signal, limits information with respect to that aspect [20, 18$^\bullet$]. Whether the response variability induced by variability in internal states increases or decreases Fisher information depends principally on two questions: first, whether the internal signal itself contains information about a, and second, whether the way it influences sensory responses is aligned with $\frac{\mathrm{d}\mathbf{f}}{\mathrm{d}a}$ [19$^{\bullet\bullet}$]. For instance, in the context of a perceptual experiment, an internal variable may represent the subject's expectations about the stimulus in the upcoming trial. If those expectations bias the sensory responses in an analogous fashion to actual sensory information, trial-to-trial variability in those expectations will induce sensory response variability in the $\frac{\mathrm{d}\mathbf{f}}{\mathrm{d}a}$ direction [16$^\bullet$]. If those expectations are based on superstition (e.g. if trials are randomized), then this internal-state-induced variability will decrease information. However, if they reflect correctly learned serial dependencies between trials, then they will increase information.

Consider the result that attention reduces covariability in the $\mathbf{f}'\mathbf{f}'^{\top}-$direction where the derivative is taken with respect to the task-relevant stimulus [35$^{\bullet\bullet}$]. Since trials were randomized in the underlying experiment, an internal variable whose value depended on previous stimulus presentations could not contain any information about the upcoming stimulus. Reducing sensory variability

with respect to such a variable would therefore *increase* information in the sensory representation in line with the behavioral improvement experimentally observed [35••].

Importantly, information is limited not by correlations themselves but by covariability, the product of correlation and individual variability. Consider Figure 5 which illustrates two different ways in which correlations in the $\mathbf{f}'-$direction can emerge. Either (panel c), by *increasing* variability that is indistinguishable from stimulus-induced variability and hence reduces information. Or (panel a), by *decreasing* variability in orthogonal directions that leave variability along $\mathbf{f}'$ and, hence, information unchanged [19••].

4 Conclusion

The influence of internal states on sensory responses can be characterized by tuning functions in analogy to the influence of the external stimulus on neural responses. Variability in those states leaves characteristic signatures in the statistics of sensory responses that can in principle be exploited to infer states and state-specific tuning functions. We suggest that analyzing sensory responses in these terms can shed new light on both their meaning and on the information they carry.

Acknowlegdements

We thank Alexander Ecker, Ruben Moreno-Bote, and Daniel Chicharro for feedback on the manuscript.

References and recommended reading

[1] Felleman, D.J. and Van Essen, D.C. (1991). Distributed hierachical processing in the primate cerebral cortex. Cerebral Cortex *1*, 1–47.

[2] Renart, A., Machens, C.K., and Alfonsorenart, A. (2014). Variability in neural activity and behavior. Current Opinion in Neurobiology *25*, 211–220.

[3] Nienborg, H., Cohen, M., and Cumming, B.G. (2012). Decision-Related Activity in Sensory Neurons : Correlations Among Neurons and with Behavior. Annual Review of Neuroscience *35*, 463–483.

[4] Georgopoulos, A.P., Kettner, R.E., and Schwartz, A.B. (1988). Primate motor cortex and free arm movements to visual targets in three- dimensional space. II. Coding of the direction of movement by a neuronal population. The Journal of Neuroscience *8*, 2928–2937.

[5] Seely, J.S., Kaufman, M.T., Ryu, S.I., Shenoy, K.V., Cunningham, J.P., and Churchland, M.M. (2016). Tensor Analysis Reveals Distinct Population Structure that Parallels the Different Computational Roles of Areas M1 and V1. PLOS Computational Biology *12*, e1005164.

[6] Mitchell, J.F., Sundberg, K.A., and Reynolds, J.H. (2009). Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4. Neuron *63*, 879–888.

[7] Cohen, M.R. and Maunsell, J.H.R. (2009). Attention improves performance primarily by reducing interneuronal correlations. Nature neuroscience *12*, 1594–600.

[8] Niell, C.M. and Stryker, M.P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. Neuron *65*, 472–479.

[9••] Cohen, M.R. and Maunsell, J.H.R. (2010). A Neuronal Population Measure of Attention Predicts Behavioral Performance on Individual Trials. The Journal of Neuroscience *30*, 15241–15253.

••Measure change in population responses as attention changes, and use this to extract the magnitude of attention on a trial-by-trial basis. Inferred attention magnitude is correlated with behavioral accuracy.

[10] Stănişor, L., Togt, C.v.d., Pennartz, C.M.A., and Roelfsema, P.R. (2013). A unified selection signal for attention and reward in primary visual cortex. PNAS *110*, 9136–9141.

[11••] Ecker, A.S., Berens, P., Cotton, R.J., Subramaniyan, M., Denfield, G.H., Cadwell, C.R., Smirnakis, S.M., Bethge, M., and Tolias, A.S. (2014). State dependence of noise correlations in macaque primary visual cortex. Neuron *82*, 235–248.

••During anesthesia, a shared source of multiplicative gain on a population changes correlation structure and inflates mean correlation estimates.

[12] Lin, I.C., Okun, M., Carandini, M., and Harris, K.D. (2015). The Nature of Shared Cortical Variability. Neuron *87*, 644–656.

[13] Schölvinck, M.L., Saleem, A.B., Benucci, A., Harris, K.D., and Carandini, M. (2015). Cortical state determines global variability and correlations in visual cortex. The Journal of Neuroscience *35*, 170–8.

[14] Nienborg, H. and Roelfsema, P.R. (2015). Belief states as a framework to explain extra-retinal influences in visual cortex. Current opinion in neurobiology *32*, 45–52.

[15] Haefner, R.M., Berkes, P., and Fiser, J. (2016). Perceptual Decision-Making as Probabilistic Inference by Neural Sampling. Neuron *90*, 649–660.

[16•] Lange, R.D. and Haefner, R.M. (2016). Inferring the brain's internal model from sensory responses in a probabilistic inference framework. bioRxiv .

•Assuming that sensory responses represent posterior beliefs, predictions are derived for the structure of feedback-induced correlations. Suggest interpretation of inferred internal states in terms of stimulus.

[17•] Tajima, C.I., Tajima, S., Koida, K., Komatsu, H., Aihara, K., and Suzuki, H. (2016). Population code dynamics in categorical perception. Nature Scientific Reports *5*, 1–13.

•Models the effect of a categorical prior, implemented through feedback, on perception in a linear probabilistic population code.

[18•] Moreno-Bote, R., Beck, J.M., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. Nature Neuroscience *17*, 1410–1417.

•Emphasize the point that response covariability in the $\mathbf{f}'\mathbf{f}'-$direction limits information, when induced by the feedforward pathway.

[19••] Ecker, A.S., Denfield, G.H., Bethge, M., and Tolias, A.S. (2016). On the structure of population activity under fluctuations in attentional state. Journal of Neuroscience *0*, 1–21.

••A theoretical study on the effect of fluctuations in attention on response correlations and their implication for Fisher Information.

[20] Abbott, L.F. and Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. Neural computation *11*, 91–101.

[21] Averbeck, B.B., Latham, P.E., and Pouget, A. (2006). Neural correlations, population coding and computation. Nature reviews. Neuroscience *7*, 358–366.

[22] Bosking, W.H. and Maunsell, J.H.R. (2011). Effects of Stimulus Direction on the Correlation between Behavior and Single Units in Area MT during a Motion Detection Task. Journal of Neuroscience *31*, 8230–8238.

[23] Haefner, R.M. (2015). A note on choice and detect probabilities in the presence of choice bias. arXiv .

[24••] 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.

••Show that choice probabilities cannot be accounted for purely by a feedforward readout and constant stimulus co-variability, implying choice-related activity in V2 contains a feedback component.

[25] Nienborg, H., Cohen, M.R., Cumming, B.G., R. Cohen, M., and Cumming, B.G. (2012). Decision-related activity in sensory neurons: correlations among neurons and with behavior. Annual review of neuroscience *35*, 463–483.

[26] 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.

[27] Pitkow, X., Liu, S., Angelaki, D.E., DeAngelis, G.C., and Pouget, A. (2015). How Can Single Sensory Neurons Predict Behavior? Neuron *87*, 411–423.

[28] Schoppik, D., Nagel, K.I., and Lisberger, S.G. (2008). Cortical Mechanisms of Smooth Eye Movements Revealed by Dynamic Covariations of Neural and Behavioral Responses. Neuron *58*, 248–260.

[29] Paninski, L., Fellows, M.R., Hatsopoulos, N.G., John, P., Wang, W., Sudre, G.P., Xu, Y., Kass, R.E., Collinger, J.L., Alan, D., et al. (2011). Spatiotemporal Tuning of Motor Cortical Neurons for Hand Position and Velocity Spatiotemporal Tuning of Motor Cortical Neurons for Hand Position and Velocity. Journal of Neurophysiology , 515–532.

[30] Haefner, R.M., Gerwinn, S., Macke, J.H., and Bethge, M. (2013). Inferring decoding strategies from choice probabilities in the presence of correlated variability. Nature neuroscience *16*, 235–42.

[31] Cohen, M.R. and Maunsell, J.H.R. (2011). Using neuronal populations to study the mechanisms underlying spatial and feature attention. Neuron *70*, 1192–204.

[32•] Cunningham, J.P. and Yu, B.M. (2014). Dimensionality reduction for large-scale neural recordings. Nature Neuroscience *17*, 1500–1509.

•Review dimensionality reduction techniques used for characterizing sensory responses.

[33] Putzky, P., Franzen, F., Bassetto, G., and Macke, J.H. (2014). A Bayesian model for identifying hierarchically organised states in neural population activity. Advances in Neural Information Processing Systems 27 , 3095–3103.

[34] Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C.E., Kepecs, A., Mainen, Z.F., Qi, X.L., Romo, R., Uchida, N., and Machens, C.K. (2016). Demixed principal component analysis of neural population data. eLife *5*, 1–36.

[35••] Rabinowitz, N.C., Goris, R.L., Cohen, M., and Simoncelli, E.P. (2015). Attention stabilizes the shared gain of V4 populations. eLife *4*.

••Modeled V4 neurons' activity as a combination of known (attentional cue, stimulus) and unknown (global gain, shared modulators) factors. Main latent factors influenced responses in the same direction as the task-relevant stimulus, implying a feedback source. modulators explain after known factors are accounted for.

[36] Purushothaman, G. and Bradley, D.C. (2005). Neural population code for fine perceptual decisions in area MT. Nature Neuroscience *8*, 99–106.

[37] Nienborg, H. and Cumming, B.G. (2007). Psychophysically measured task strategy for disparity discrimination is reflected in V2 neurons. Nature neuroscience *10*, 1608–14.

[38••] Cohen, M.R. and Newsome, W.T. (2008). Context-Dependent Changes in Functional Circuitry in Visual Area MT. Neuron *60*, 162–173.

••First to show task-dependence of response correlations in MT neurons.

[39••] Bondy, A.G. and Cumming, B.G. (2016). Feedback Dynamics Determine the Structure of Spike-Count Correlation in Visual Cortex. bioRxiv , 1–41.

••Fit full noise correlation matrices to macaque V1 neurons in changing task contexts. The full matrix is dominated by a task-dependent component. It is predominantly rank-1 and approximately proportional to $\frac{\mathrm{d}\mathbf{f}}{\mathrm{d}x}$ (Box 1).

[40] Markov, N.T., Ercsey-Ravasz, M., Van Essen, D.C., Knoblauch, K., Toroczkai, Z., and Kennedy, H. (2013). Cortical high-density counterstream architectures. Science *342*, 1238406.

[41] Sherman, S.M. (2016). Thalamus plays a central role in ongoing cortical functioning. Nature neuroscience *16*, 533–41.

[42] Briggs, F., Mangun, G.R., and Usrey, W.M. (2013). Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits. Nature *499*, 476–80.

[43] Wimmer, K., Compte, A., Roxin, A., Peixoto, D., Renart, A., and Rocha, J.D. (2015). The dynamics of sensory integration in a hierarchical network explains choice probabilities in MT. Nature Communications *6*, 1–13.

[44] Moore, T. and Zirnsak, M. (2017). Neural Mechanisms of Selective Visual Attention. Annual Review of Psychology *68*, 47–72.

[45] Ahissar, M. and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. Trends in Cognitive Sciences *8*, 457–464.