Dynamical processing of orientation precision in the primary visual cortex

Hugo J. Ladret[1, 2, ✉], **Nelson Cortes**[2], **Lamyae Ikan**[2], **Frédéric Chavane**[1], **Christian Casanova**[2], **and Laurent U. Perrinet**[1]

[1] Institut de Neurosciences de la Timone, UMR 7289, CNRS and Aix-Marseille Université, Marseille, France
[2]School of Optometry, Université de Montréal, Montréal, QC H3C 3J7, Canada

The primary visual cortex (V1) processes complex mixtures of orientations to build neural representations of our visual environment. It remains unclear how V1 adapts to the highly volatile distributions of orientations found in natural images. We used naturalistic stimuli and measured the response of V1 neurons to orientation distributions of varying bandwidth. Although broad distributions decreased single neuron tuning, a neurally plausible decoder could robustly retrieve the orientations of stimuli from the population activity. Furthermore, we showed that V1 co-encodes orientation and its precision, which enhances population discriminatory performances. This internal representation is mediated by temporally distinct neural codes, supporting a precision-based description of the neuronal message passing in the visual cortex.

orientation | decoding | precision | predictive coding | V1
Correspondence: ***hugo.ladret@univ-amu.fr***

Introduction

Selectivity to the orientation of stimuli is an archetypal feature of neurons in the primary visual cortex (V1) (1). In more than 60 years of investigations, low-complexity stimuli such as gratings have uncovered many orientation-tuned filter-like properties in V1 (2). These stimuli are, however, intrinsically limited in their relevance to study the rich cortical dynamics (3) involved in the vision of natural images (4). Indeed, the complex content of our everyday visual scenery (5) enforces sparse and efficient population codes in V1 (6–8). Thus, any integrative understanding of how the early visual system works has to account for the behaviour of V1 in its ecological visual environment (9). While it is possible to directly probe V1 with natural images (10), synthetic "natural-like" stimuli are often preferred (11). This approach offers a compromise between the desire for keeping a fine control over the experiments and the need for complex, naturalistic stimuli to study ecological brain mechanisms. Moreover, these synthetic stimuli allow the isolation of relevant visual parameters by removing correlations between the multiple perceptual dimensions composing a natural image (12–15).

The underlying structure of natural images and natural-like stimuli can be represented as a distribution of orientations (Supplementary Figure 1). The amplitude of such distributions corresponds to the contrast of the image, whose impact on orientation selectivity has been thoroughly investigated (16). However, the effect of the broadness of this distribution (i.e. the orientation bandwidth), remains unexplicit (14, 17). Broader orientation bandwidths reduce the salience and the precision of orientation inputs. This is a central challenge which V1 must constantly address, as the distribution of orientations in natural images is highly heterogeneous. Current models of the influence of orientation bandwidth come from the field of texture perception (18). Intuitively, psychophysics experiments have shown that increased broadness in orientation distributions worsens orientation discrimination performances (19–21), which can be explained by computational models of intracortical dynamics (22, 23). Despite these converging observations, there has been very few investigations in the neural correlates of the processing of orientation bandwidth.

Here, we used naturalistic stimuli of controlled orientation distribution to study the encoding of bandwidths of orientations in V1. Using a biologically plausible neuronal decoder (24), we found that stimuli orientation could be robustly inferred from the population activity at multiple broadness of input bandwidth. While this variable is not encoded explicitly in the activity of individual V1 neurons, we showed that orientation and its bandwidth are co-encoded, which improves the orientation discriminatory performances of a neural population. Moreover, the bandwidth of an orientation is correlated with its processing time, which stems from temporally separated neural representations of orientation bandwidths for the same neural network. Orientations of narrow bandwidths are encoded in the population activity right after the onset of the stimuli. This representation then fades away to make place for a broader bandwidth neural code, which can peak up to several hundred milliseconds later. These results suggest that the precision of orientation features, in the form of their bandwidth, is encoded in V1, which conveys crucial implications for theories requiring the encoding of the precision of sensory variables, such as predictive coding (25).

Results

Single neuron modulation by orientation bandwidth. Single-unit activity of 254 V1 neurons was recorded in three anesthetized cats. Orientation selectivity was measured using band-pass filtered white noise images called Motion Clouds (26), whose underlying generative framework allowed to precisely control the orientation content of each visual stimulus. Specifically, we varied two parameters of Motion Clouds: their main orientation θ and the bandwidth of the orientation distribution B_θ (Figure 1a). Like natural images, Motion Clouds are complex mixtures of orientations, but their stationarity in the spatial domain removes any poten-

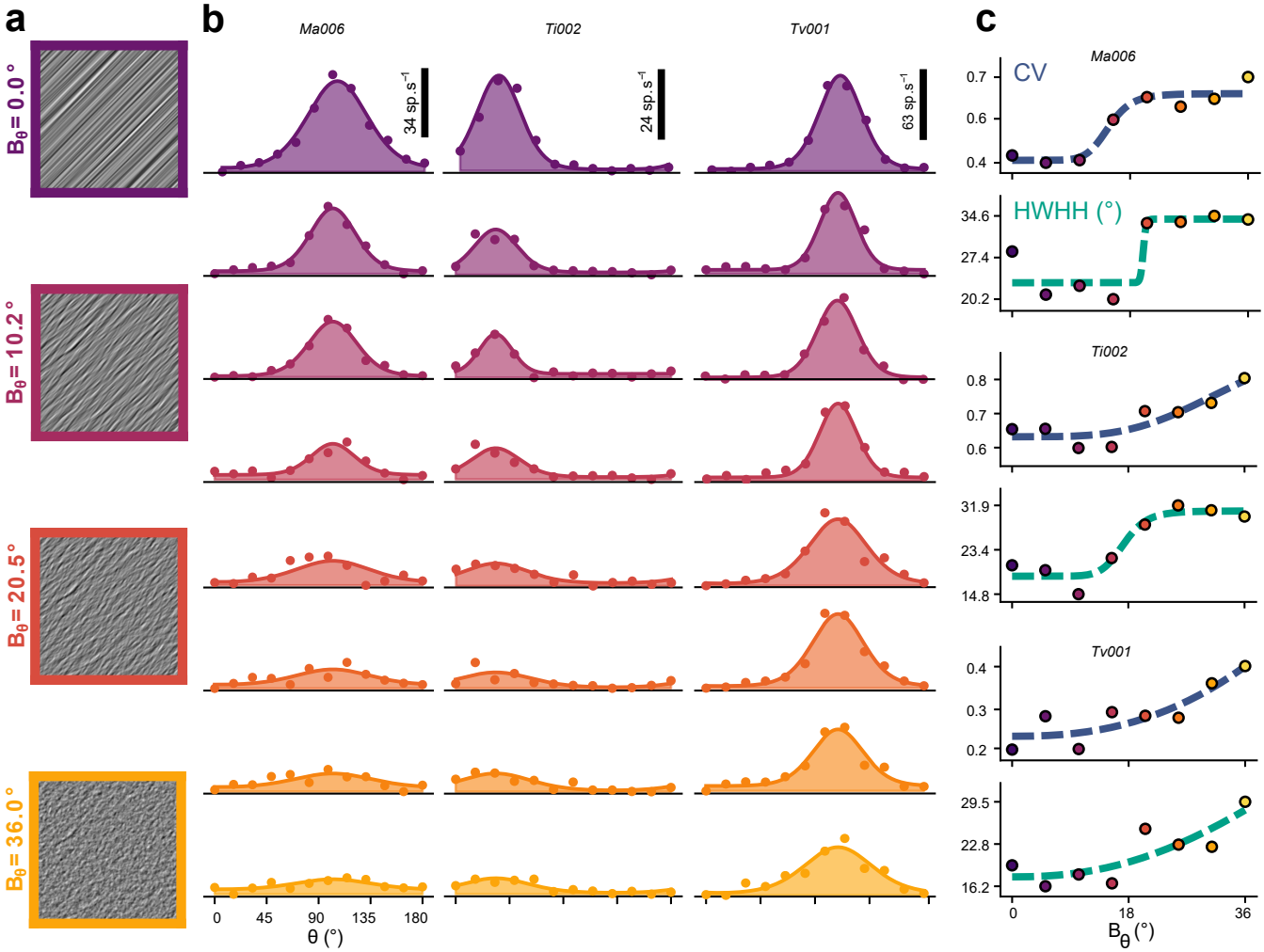


Figure 1. Band-pass filtered white noise stimuli and single-neuron responses. (**a**) Examples of Motion Clouds, oriented at 45° relative to the horizontal axis. Neuronal responses were measured with orthogonally drifting Motion Clouds of 12 different orientations θ and 8 orientation bandwidth B_θ. (**b**) Tuning curves of three examples neurons stimulated with Motion Clouds of increasing B_θ, i.e. increasing orientation bandwidth (from top to bottom). Colored points represent the mean firing rate across trials (baseline subtracted, 300 ms average), with lines indicating a fitted von Mises function. (**c**) Changes in orientation tuning were measured by the Half-Width at Half-Height (HWHH) and Circular Variance (CV) as a function of B_θ and fitted with Naka-Rushton functions (dashed curves). Additional examples are shown in Supplementary Figure 2.

tial second-order correlation, and allows to isolate the effect of B_θ on orientation selectivity (Supplementary Figure 1). Moreover, they conform to the statistics of natural images, namely the $1/f^2$ power spectrum distribution (6).

We measured the orientation selectivity of V1 neurons to Motion Clouds for 12 evenly spaced values of θ (steps of 15°) and 8 values of B_θ (steps of 5.1°, starting at 0°). As B_θ increased, the vast majority of tuning curves remained centered on the same preferred orientation (98% units, $p < 0.05$, Wilcoxon signed-rank test) but diminished in peak amplitude (94% units, $p < 0.05$, Wilcoxon signed-rank test, 73.1% mean amplitude decrease). Only 26% neurons were still tuned when orientation bandwidth reached $B_\theta = 36.0°$. As such, the common selectivity modulation caused by an increasing orientation bandwidth B_θ was a broadening of tuning, which unfolded heterogeneously between neurons, as illustrated by three example tuning curves (Figure 1b). Namely, neurons Ma006 and Ti002 were no longer tuned to stimuli of $B_\theta = 36.0°$ ($p = 0.24$ and 0.07, respectively, Wilcoxon signed-rank test, firing rate of preferred vs orthogonal orientation), contrary to neuron Tv001 which remained orientation selective ($p = 10^{-6}$). The variations of tuning broadness were more precisely assessed by measuring both the global tuning of the raw data (Circular variance, CV) and the local strength of tuning around the peak of a fitted von Mises function (Half-Width at Half-Height, HWHH). The step-like changes observed (Figure 1c) were fitted with a Naka-Rushton function (27) (Supplementary Figure 3), allowing the estimation of two descriptive parameters of the bandwidth-response function. Firstly, the bandwidth value B_{50} corresponds to the B_θ at which a neuron transitioned towards a more broadly tuned state. The second one, the slope n, indicates the steepness of the variations of tuning as B_θ increases. n was correlated to the tuning width at $B_\theta = 0.0°$ ($p_{\text{HWHH}} = 0.048$, Spearman's $\text{R}_{\text{HWHH}} = 0.12$, $p_{\text{CV}} = 10^{-6}$, Spearman's $\text{R}_{\text{CV}} = 0.33$), that is, broadly tuned neurons underwent the largest tuning modulations as B_θ increased. No other significant relationship was found, except for the n/B_{50} correlation produced by the Naka-Rushton fitting procedure (the optimal fit of a neuron whose B_{50} is past

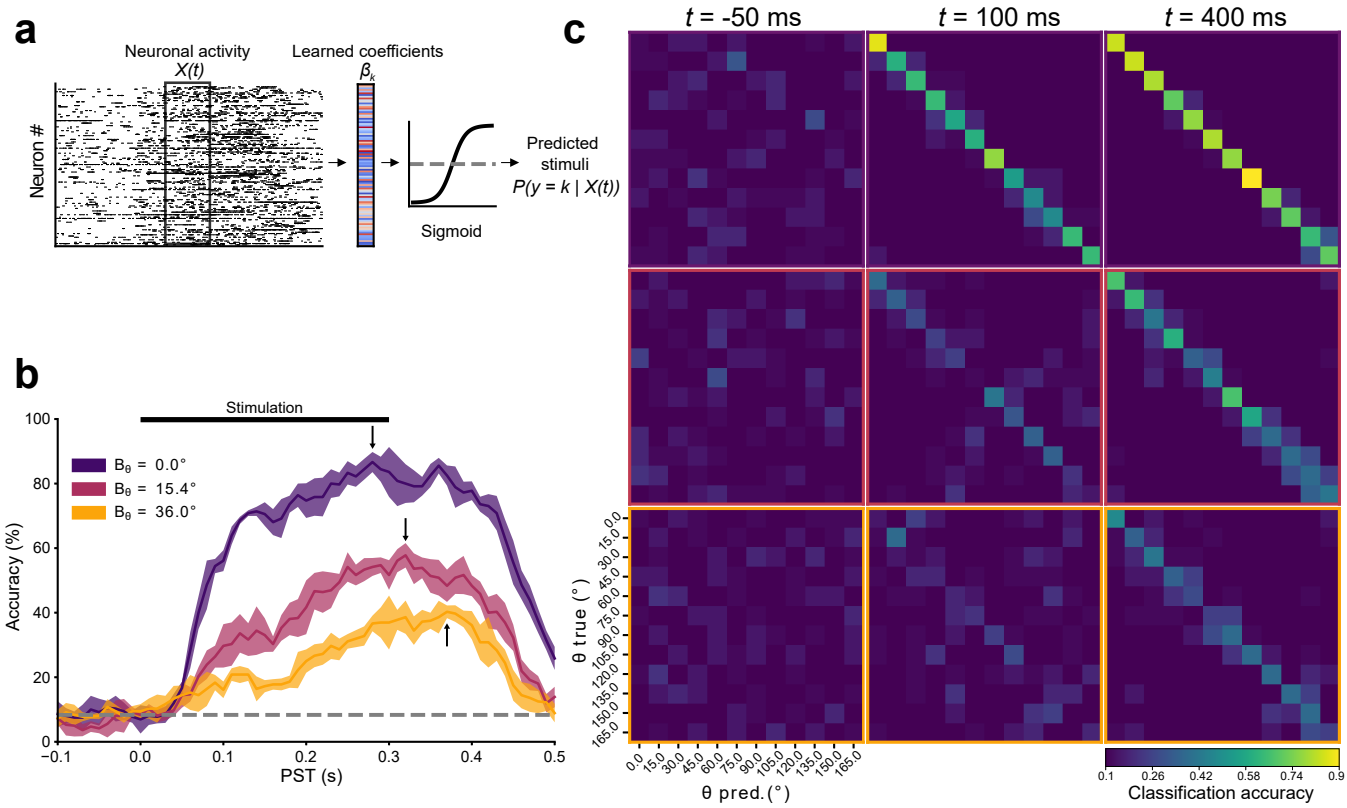


Figure 2. Decoding orientation from population activity. (**a**) Illustration of the decoding model. For each trial, spikes from all single units in a given time window were pooled in a single neural activity vector $X(t)$. A multinomial logistic classifier predicts the identity k of the stimulus in this trial (here, θ_{true}) by estimating the probability for different hypothesis y (here denoted $\theta(t)_{\text{pred}}$). This is achieved by learning a weight vector β_k which maximizes the accuracy of the prediction $Pr(y = k|X(t)) \approx \sigma(\beta_k^T X(t))$, where σ is the sigmoid function. This accuracy serves as a proxy for the discriminability of the population's neural activity within feature space. (**b**) Time course around the peri-stimulus time (PST) of the accuracy of three independent decoders trained to predict the orientation θ of Motion Clouds for a given orientation bandwidth B_θ. Solid dark line represent the mean accuracy of a 5-fold cross validation and filled contour the SD. Decoding at chance level (here, $\frac{1}{12}$) is represented by a gray dashed line. Arrows indicate the time at which maximum average accuracy was reached. (**c**) Confusion matrices of the decoders at three time points around stimulation time. Values inside the diagonals of each matrix represent the instances of correct decoding, that is, where $\theta(t)_{\text{pred}} = \theta_{\text{true}}$. Each row of matrices belongs to the decoder with the same color code in (b).

the maximum measured B_θ is a linear function). Overall, the population's parameters were heterogeneous (Supplementary Figure 4), suggesting a diversity of tuning modulations between neurons would provide better orientation selectivity capacity between multiple orientation bandwidths (14).

Decoding orientation from population activity. The variety of tuning modulations observed in single neurons would suggest that a form of population code is required to decode orientation through all values of B_θ. We used a neuronal decoding method to see if the parameters of the Motion Clouds could be read out from the population activity. Briefly, we trained a multinomial logistic regression classifier, a probabilistic model that classifies data with multiple outcomes (see Material and Methods). Here, this classifier was fed the firing rate of all recorded neurons in a given time window and learned to predict the identity of a stimulus eliciting the neural activity (Figure 2a). To decode orientation θ, the dataset of trials were split by B_θ such that 8 independent orientation-bandwidth specific decoders were learned. These were then able to predict the correct Motion Clouds' θ well above the chance level of 1 out of 12 θ. The temporal evolution of decoding performance for these decoders (Figure 2b) shows that the maximum is reached faster for stimuli with narrower bandwidths (270 and 370 ms post-stimulation for $B_\theta = 0.0°$ and $B_\theta = 36.0°$, respectively). The activity is sustained up to 150 ms after the stimulation could serve to decode the orientation, even more so for broad B_θ Motion Clouds. A more detailed description of these decoders' performances is given by confusion matrices (Figure 2c), which represent the accuracy of the decoder for all possible combinations of true and predicted stimuli (see Materials and Methods). The maximal diagonalization of the $B_\theta = 0°$ decoder is close to the diagonalization of the $B_\theta = 36.0°$ decoder (12/12 versus 11/12 classes significantly decoded above chance level, respectively, $p < 0.05$, 1000 permutations test). Thus, orientation could be retrieved above chance level from the population activity. The short delay required to perform decoding above chance level is congruent with the feedforward processing latency of V1 (24), while the increased time required to reach maximum accuracy, especially for broad bandwidth orientations, would hint at the presence of a slower mechanism at these precision levels.

Decoding orientation bandwidth. As the decoding of orientation was impacted by orientation bandwidth, we sought to understand if B_θ was also encoded in the population activity. A decoder trained to retrieve the B_θ of stimuli (Figure 3a)

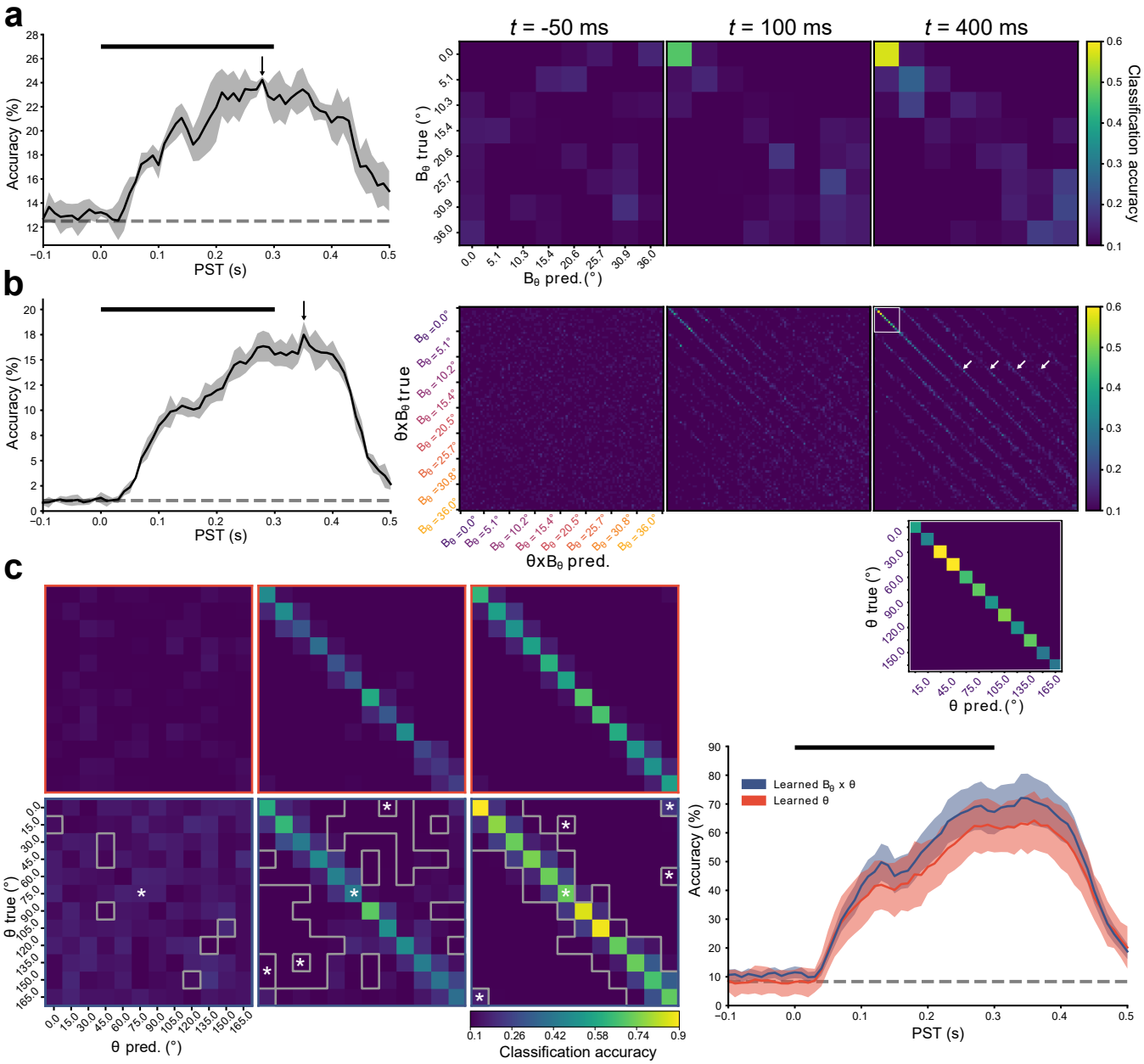


Figure 3. Decoding orientation bandwidth. (**a**) Left: Time course of a decoder trained to retrieve the orientation bandwidth (B_θ) of Motion Clouds. Solid dark line represent the mean accuracy of a 5-fold cross validation and filled contour the SD. Arrows indicate the point at which maximum accuracy was reached. Right: Confusion matrices of the decoder. (**b**) Left: Time course of a decoder predicting both the orientation (θ) and orientation bandwidth (B_θ). Arrows indicate the point at which maximum accuracy was reached. Right: Confusion matrices of the decoder. In each section labeled by a given B_θ, the local matrix organization is identical as in Figure 2, as represented in the white inset. White arrows indicate the B_θ erroneous, θ correct diagonals parallel to the central identity line. (**c**) Left, top: confusion matrices of a decoder trained to retrieve orientation θ without prior knowledge of B_θ. Left, bottom: confusion matrices of the $B_\theta \times \theta$ decoder (fig. 3b), marginalized over B_θ to show θ decoding accuracies. Grey contoured regions congruent to a white asterisk indicate a significant difference between the two decoders (1000 permutations test, n=6, threshold: p < 0.05). Right: time courses of the marginalized decoder (blue, bottom confusion matrix row) and orientation-only decoder (red, top confusion matrix row).

performed at a maximum accuracy of twice the chance level in 280 ms. However, the confusion matrix indicates that the predictions were essentially binary, discriminating between narrow and broad B_θ ($B_\theta = [0.0°; 5.1°; 36.0°]$ significantly decoded above change level, $p < 0.05$, 1000 permutations test). This dyadic decoding is reminiscent of the step-like changes observed in tuning curves (Figure 1c), but also of the distributions of the tuning steepness in the population (Supplementary Figure 4b). Altogether, a binary estimation of the orientation's bandwidth is available independently of the orientation itself, but a fine representation of bandwidth, especially for intermediate B_θ, cannot be retrieved accurately as a standalone variable using the present decoding method.

Nevertheless, orientation bandwidth could also be encoded in V1 as a two-dimensional variable, consisting of the Cartesian product of orientation and bandwidth. Such a $B_\theta \times \theta$ decoder was trained and could retrieve the identity of Motion Clouds with a maximum accuracy of about 16 times the chance level $\frac{1}{96}$ in 350 ms (Figure 3b). This delay to maximum accuracy is similar to the delay of broad B_θ orientation decoders (Figure 2b), suggesting that stimulus' identity can only be properly inferred when both narrow and broad B_θ

stimuli have been decoded. In line with the previous θ decoders (Figure 2b), the confusion matrix of this $B_\theta \times \theta$ decoder showed greater accuracy for patterns of narrow bandwidth (Figure 2b, upper left corner, 44% average θ decoding for $B_\theta = 0.0°$ and 12% for $B_\theta = 36.0°$). Only 26% of stimuli were accurately predicted to a significant degree due to secondary, parallel diagonals in the confusion matrices (Figure 3b, right, arrows), which are instances where the decoder correctly predicted θ with no exact certainty on the associated B_θ. Since the addition of B_θ worsens the decoding of θ, what advantage is there to co-encode bandwidth? To answer this question, we compared the performances of two orientation decoders (Figure 3c). The first learned only θ without any prior knowledge on B_θ while another was the marginalization of the $B_\theta \times \theta$ decoder over B_θ (Figure 3b). This is a simple re-structuring operation which "folded" the 96 class confusion matrix over its columns and rows to obtain an orientation-only output. When marginalized, the $B_\theta \times \theta$ decoder performed better than the B_θ-agnostic decoder (12/12 classes significant, $p < 0.05$, 1000 permutations test). Prior knowledge on the orientation bandwidth associated with an oriented input thus yielded better performances of orientation decoding.

The visualization of the decoder's weight gave some indication as to how this dual $B_\theta \times \theta$ code was implemented by single neurons. The decoder learned for each neuron and at each time step a vector β_k of K elements. Aligning all the vectors through time created a coefficient map that allowed to visualize the contribution by individual neurons to the population code in the $B_\theta \times \theta$ decoding paradigm. Repeated values through B_θ in each vector (Figure 4a, red areas in each dashed sections) correspond to a "vote" for the preferred orientation of the neuron, along with ambiguity over multiple B_θ. Similar to the behavior of tuning curves (Figure 1b), the coefficient maps showed a loss of orientation selective decoding coefficients as B_θ increased for some neurons (coefficient maps of neurons Tc001 and Th002), while others remained tuned with an increasing width and time delay (neuron Th004). While the tuning curves were linked to the neuron's receptive field, the coefficient maps here are comparable to an "emitting field", i.e. to the nature of the message passed by a neuron to the population. The presence of repeated peaks raised one important question: if the B_θ identity of orientation θ is ambiguous in single neurons (repeated values, Figure 4a) and in the population (secondary diagonals, Figure 3b), how can V1 retrieve the $B_\theta \times \theta$ identity of a stimulus? If a neuron encodes one θ over multiple B_θ, a possible disambiguation of the stimulus' identity comes from the temporality of the B_θ decoding. The specific time course of bandwidth decoding creates a lagged onset for broad bandwidth orientations, particularly visible in the coefficient map of neuron Th004. This is for instance reflected in the decoder's total pool of coefficients, in which the narrower half of measured bandwidths ($B_\theta < 18°$) stimuli are best encoded during the first 200 ms post-stimuli, transitioning to a better representation for the broader half ($B_\theta > 18°$) afterwards (Figure 4b).

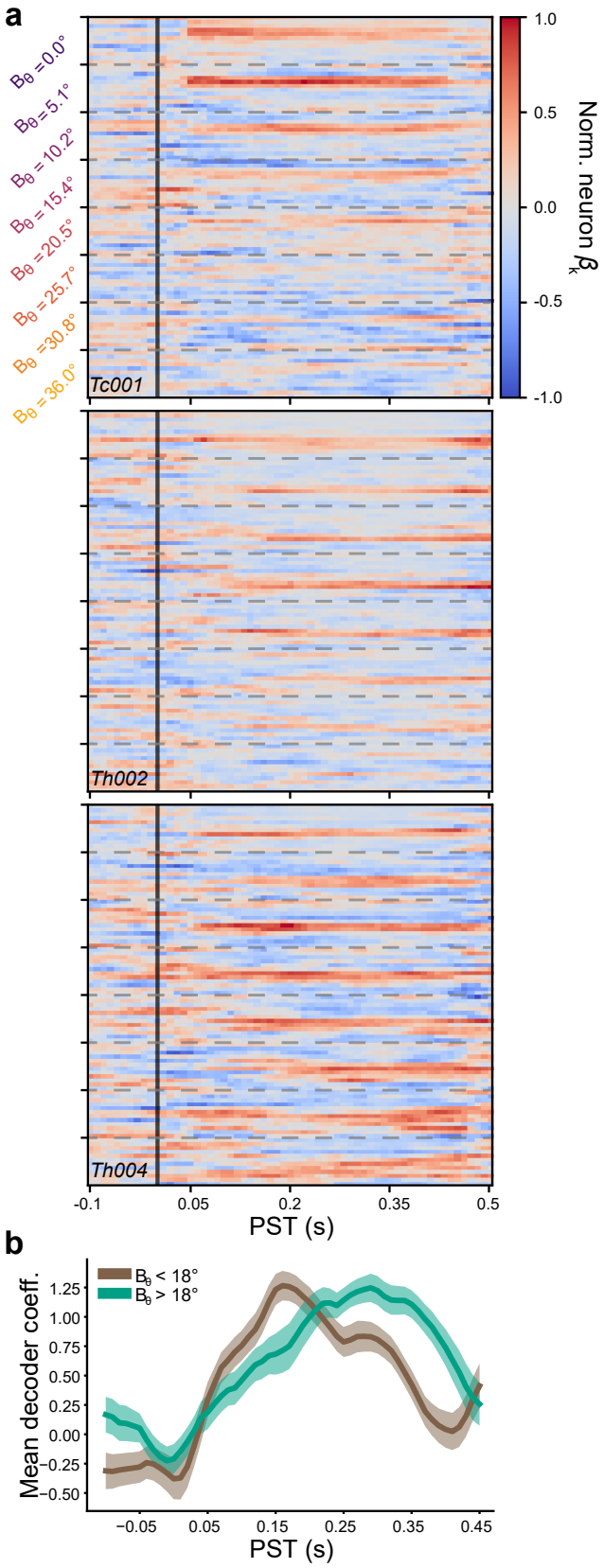


Figure 4. Neuronal contributions to population decoding. (**a**) Coefficient maps around the peri-stimulus time (PST) of three neurons. (**b**) Time course of the total coefficients of the $B_\theta \times \theta$ decoder, averaged on a 50 ms window. Darker lines indicate mean values and filled contour SEM.

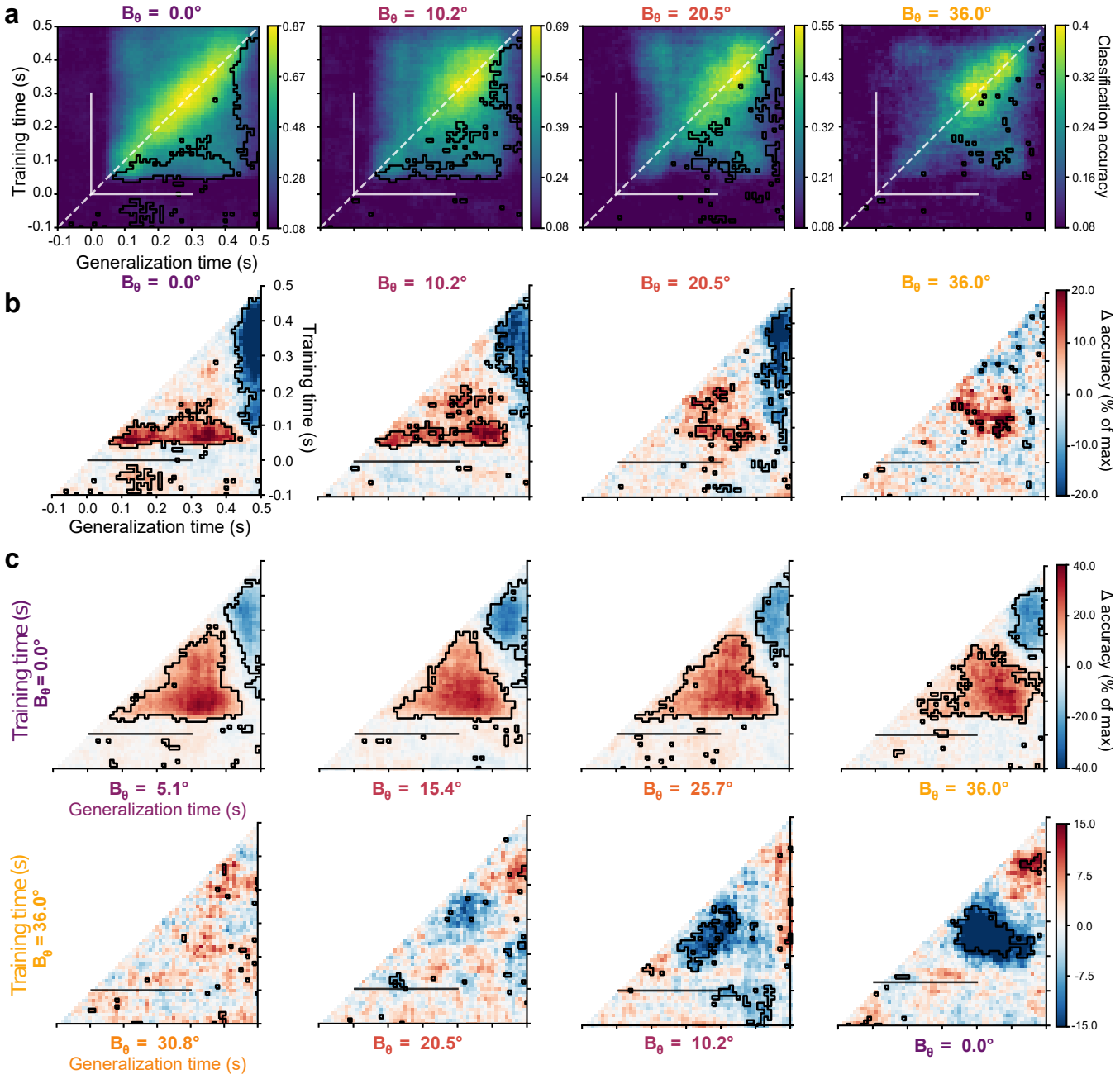


Figure 5. Temporal generalization. (**a**) Temporal generalization matrices of four independent orientation decoders trained for different bandwidths. Black contoured regions indicate significant difference (1000 permutations test, n=6, threshold $p < 0.05$) compared to the symmetrical region with respect to the time identity line (white dotted line). White lines represent stimulation duration. (**b**) Asymmetry of the temporal generalization matrices. Δ accuracy is the difference of accuracy between the upper and lower halves of the matrices shown in (**a**), expressed as a percentage of each matrix's maximum accuracy. Black lines represent stimulation duration. (**c**) Top row: Asymmetry of the cross-conditional temporal generalization matrices shown in Supplementary Figure 8b) training a decoder on the narrowest orientation bandwidth and generalizing it onto broadening B_θ (from left to right). Bottom: Generalizing from broadest to narrowest orientation bandwidths (from left to right).

Temporal dynamics of bandwidth decoding. As the decoding of orientation and bandwidth evolved through trial time, we investigated the detailed dynamics of the neural code by using temporal generalization (28) (see Materials and Methods). Briefly, a decoder was trained and tested at two different time points, $train_t$ and gen_t, with the underlying hypothesis that a successful temporal generalization indicates a similar neural code at both training and testing times. Generalization points where $train_t = gen_t$ (on the white dashed line, Figure 5a) are the exact same measurement as the time course of accuracy shown previously (Figure 2b). As the decoder's accuracy represents the level of discriminability of the population's activity within feature space, measuring the performance of the decoder when $train_t \neq gen_t$ allowed us to assess the unraveling of the neural code in time. The upper halves of the temporal generalization matrices above the time identity line correspond to generalizations where $train_t \geq gen_t$, i.e. generalizing backward in trial time, while generalizations below the time identity line, where $train_t \leq gen_t$, are generalizations forward in trial time. Hence, sub-

tracting the upper half to the lower half of each matrix measures the asymmetry around the time identity line (contoured clusters, Figure 5, Supplementary Figure 8) and gives an indication of the direction of the neural code in the time dimension. Positive asymmetry (red clusters) indicates a better generalization forward in trial time, while negative asymmetry (blue clusters) indicates a better backward generalization. For instance, the temporal generalization matrices of narrow B_θ-specific orientation decoders showed an asymmetry around the time identity line (Figure 5b), meaning that upper and lower regions of the matrix were significantly different (44% significantly asymmetric transpositions, $B_\theta = 0.0°$, $p < 0.05$, 1000 permutations test). In other terms, training a narrow B_θ-specific orientation decoder and transposing it further in time within the trial (lower half of each generalization matrix) is not identical to training and transposing it backwards in trial time (upper half). Notably, this asymmetry is positive when $train_t \approx 100$ms and $gen_t \geq 100$ms, implying that narrow B_θ-specific orientation decoders can use early trial activity to infer the identity of the stimulus during the rest of the trial, hence that narrow B_θ information is encoded in the early transient population activity. Further, the asymmetry became negative for $gen_t \geq 350$ms, implying that the late activity does not contain any information relevant to the decoding of earlier narrow B_θ stimuli. For generalization matrices of broad B_θ orientation decoders, the proportion of significantly asymmetrical transpositions is almost null : the transposition matrix is symmetric, i.e. the late neural code can be transposed backward or forward in time (1% significantly asymmetric transposition, $B_\theta = 36.0°$, $p < 0.05$, 1000 permutations test). Taken together, the asymmetry of the $B_\theta = 0.0°$ generalization matrix shows that the early population activity is suited to decode narrow B_θ, whereas the symmetry of the $B_\theta = 36.0°$ matrix implies that the late neural code emerges after the narrow representation, culminating late in trial time.

To understand the extent to which each B_θ-specific code extends temporally, matrices of cross-conditional temporal generalization were produced, where training and testing bins came from different B_θ (Figure 5c, full matrices shown in Supplementary Figure 8b). The asymmetry previously observed for narrow B_θ generalization remained marked when transposing the $B_\theta = 0.0°$ orientation decoder to broader B_θ stimuli (55% significantly asymmetric transpositions from $B_\theta = 0.0°$ to $B_\theta = 36.0°$, $p < 0.05$, 1000 permutations test). The fact that the positive asymmetry remains confined to the early $train_t$ while negative asymmetry is still present in late $train_t$ shows that a narrow B_θ code is always present at the transient of the stimulation and stays similar across time points, no matter the broadness of the input. The transposition of the decoder trained for $B_\theta = 36.0°$ displayed a progressive asymmetrization when transposed to patterns of narrower bandwidth (20% significantly asymmetric transpositions from $B_\theta = 0.0°$ to $B_\theta = 36.0°$, $p < 0.05$, 1000 permutations test). Namely, a pattern of negative asymmetry was present for $train_t \approx 150$ms and $gen_t \geq 250$ms, which corresponds to the region of non-significant asymmetry in

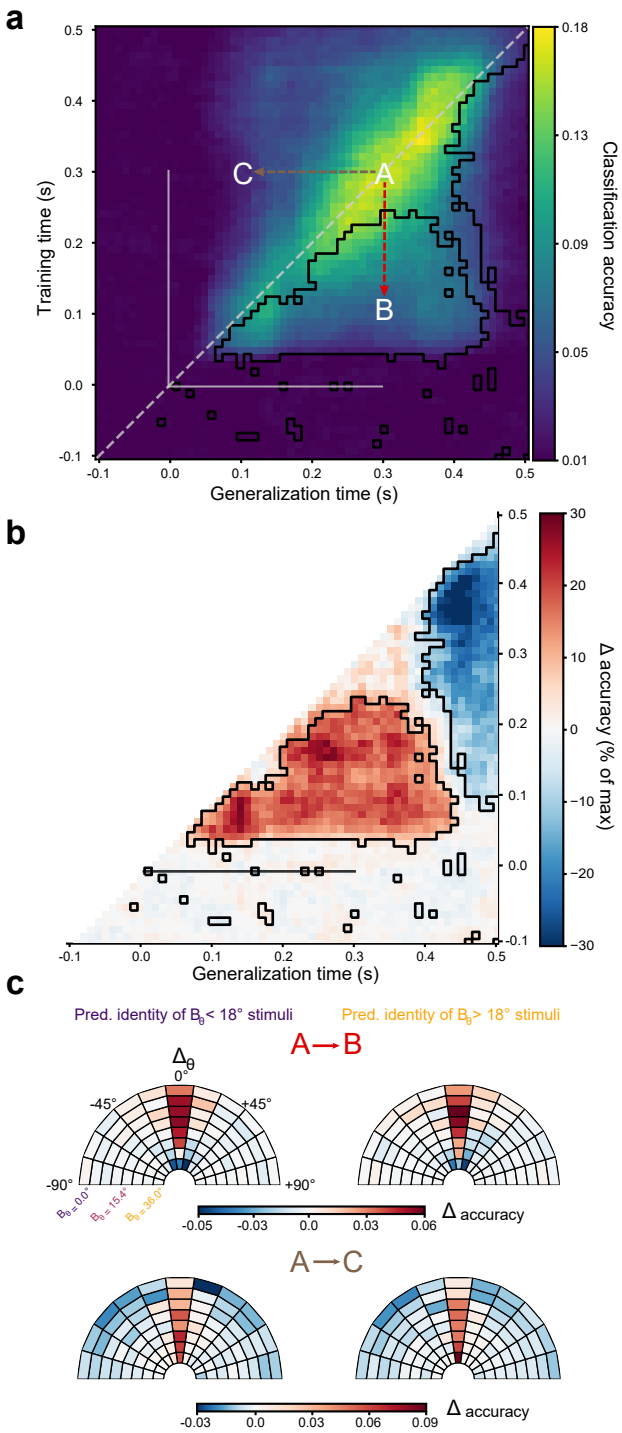


Figure 6. Temporal generalization of the orientation and orientation bandwidth ($B_\theta \times \theta$) decoder. (**a**) Temporal generalization matrix of the decoder. Three time points of interests are labeled A ($train_t = 300 = gen_t = 300$ ms), B ($train_t = 100; gen_t = 300$ ms) and C ($train_t = 300; gen_t = 100$ ms). Gray lines represent stimulation duration. Contoured regions indicate significant difference (1000 permutations test, n=6, threshold p < 0.05) compared to the symmetrical region with respect to the time identity line (dashed gray). (**b**) Asymmetry of the temporal generalization matrix. Black line represent stimulation duration and black contours regions are of significant asymmetry, as previously. (**c**) Polar representation of the confusion matrices differences. Stimuli were re-aligned onto their main orientation ($\Delta_\theta = 0°$) to measure the decoding's broadness in the orientation domain (see also Supplementary Figure 7).

the $B_\theta = 0.0°$ iso-conditional transposition (Figure 5b). This region in which the broad bandwidth decoder cannot decipher narrower bandwidth population activity corresponds to the transition period in the decoder's coefficients where narrow B_θ code starts receding, but population activity encoding broad B_θ has yet to arise (Figure 4b). As such, the neural codes for narrow and broad bandwidth seem to follow two different temporal gradients, in which the transient activity evoked by a stimulus encodes a narrow bandwidth activity, which then progressively disappears in favor of broad bandwidth activity.

While the generalization of orientation decoders allowed to separate the relative timing of the different bandwidth decoders, we have stated that co-dependency on B_θ and θ improves the population's code in V1 (Figure 3c). Hence, to understand how the two dynamics of narrow and broad bandwidth codes are combined into a single population representation, we carried out the temporal generalization of the $B_\theta \times \theta$ decoder (Figure 6a). The characteristic asymmetry of the previous narrow B_θ matrices was observed (78% significantly asymmetric transpositions, $p < 0.05$, 1000 permutations test). Coherently, the asymmetry showed a first positive region, corresponding to the previously observed preference for time-forward generalization of narrow B_θ (black contoured red region, Figure 6b), before transitioning to a negative asymmetry for the putative time-backward, broad B_θ code. The overall $B_\theta \times \theta$ population code retains the specificity of the bandwidth-dependent dynamics.

A more specific representation of the preference of the $B_\theta \times \theta$ decoder towards certain bandwidths is obtained by marginalizing the confusion matrices over B_θ (Supplementary Figure 7). By computing the difference of accuracy between two such marginalized matrices, a detailed exemplification of the decoding is obtained (Figure 6c). In this polar representation, we show the difference of accuracy as a function of the difference between the predicted and true orientation (Δ_θ) and of the bandwidth B_θ. A difference between a transposition point on the identity line and one on the lower half of the transposition matrix (Figure 6c, A to B difference, upper row), which is a difference forward in the time within the trial, shows that a narrow B_θ stimulus is preferentially decoded, regardless of the actual stimulus identity. Conversely, the difference of accuracies between a point on the identity line and one on the upper half of the transposition matrix (Figure 6c, A to C difference, lower row), i.e. a difference backward in trial time, shows a better decoding of broad bandwidth stimulus. The difference between A to B and A to C was asymmetric : while the first one shows a preference for narrow B_θ in forward transpositions, the second is suited to broader B_θ but with a greater Δ_{accuracy}.Both observations comfort the previously established hypothesis of the two different temporal gradients. In the orientation domain, Δ_θ shows a highly selective modulation on the orientation of the stimulus, with furthest orientation deviations on the order of one step in stimulation parameters (15°). The $B_\theta \times \theta$ population code retains the characteristic asymmetry of the B_θ-specific codes, while also staying well tuned to orientation. Information on the orientation bandwidth emerges differentially through time, which could serve to signal the precision of the sensory input and reduce ambiguity in the $B_\theta \times \theta$ space.

Discussion

The bandwidth of orientation distributions impacts orientation selectivity and we have sought to understand how V1 processes this critical visual feature. Using naturalistic stimuli, we showed that non-linearities of single neurons support a population activity that co-encodes orientation and bandwidth through differential dynamical processes.

Few studies have investigated the neural correlates of orientation bandwidth. Recently, Goris *et al.* (14) used stimuli similar to Motion Clouds to investigate the function of tuning diversity in V1 and V2, reporting that heterogeneously tuned populations are better suited to encode the mixtures of orientations found in natural images. Consistent with their findings, we reported a variety of single-neuron modulations by B_θ and a steepness-tuning correlation evocative of a relationship between tuning width and stimulus (29). Therefore, an heterogeneously tuned population, both in θ and B_θ, seems to be suited to decode orientation mixtures of varying bandwidth. This type of population activity is known to be pivotal in the responses to natural images (30, 31). The decoder used presently can be interpreted as a linear-nonlinear neuron (24, 32) performing a readout of V1's activity. For narrow B_θ stimuli, above chance level readout was performed in a timing coherent with feedforward processing, while the delay required to infer the identity of stimuli with maximum accuracy, especially for broader B_θ, would suggest the involvement of a second, non-feedforward neuronal pathway.

One candidate for this process could be the local cortical connectivity in V1. Indeed, recurrent connectivity implements complex computations (33) and maintains tuning to features that cannot be encoded by single neurons (34), making it relevant in processing stimuli such as Motion Clouds. In this hypothesis, the increased delay to reach maximum accuracy for broad bandwidth stimuli would stem from multiple iterations of computations implemented by recurrent interactions, which would be coherent with the speed of slowly-conducted horizontal waves in the orientation map (35). Likewise, stimuli of broad orientation bandwidth should recruit more neighboring columns of different preferred orientations, whose inhibitory interactions in the orientation domain would explain the decrease in single neurons response, as observed with plaid stimuli (36). In line with this idea, the diversity of single-neurons modulations reflects that the heterogeneity in the recurrent synaptic connectivity can sustain a resilient tuning to orientation selectivity in V1 (37). A second suitable origin of the precision modulation might come from extrastriate feedback, namely in the form of inference on the state of V1 (38). In terms of a predictive framework (25), parallel processing of feedback predictions on sensory states conflicts with feedforward activity. The difference between predicted and actual sensory states is called the prediction error, and serves to update the predictions. This update rule

is modulated by the precision of the prediction, such that high precisions expectations cause a greater weighting on prediction adjustment. As orientation bandwidth and precision, in the form of the inverse variance, are tightly correlated, stimuli of varying bandwidths are a useful investigation tool to understand the neuronal implementation of the precision weighting of prediction errors. This feedback-centric interpretation of our results would imply that two low and high precision codes co-exists temporally and spatially, corresponding respectively to the feedforward sensory input and the feedback precision-weighted prediction. In anesthetized cats, feedback from extrastriate areas modulates the gain of V1 responses without effect on the orientation preference (39) but provides contextual modulations of V1 (40), which is a suitable synonym for precision weighting (41). Quite possibly, the interaction between recurrent and feedback information (42) is capital to the precision modulation of message passing in the brain. Some consideration should also be given to transthalamic pathways, namely those involving the pulvinar nucleus, whose reciprocal connectivity with the visual hierarchy modulates cortical response functions (43) and is theorized to finesse the precision weighting of visual information (41).

We found no evidence for an explicit neural code for precision that would be independent of orientation, at least using our current decoding paradigm. However, the multi-feature tuning observed between θ and B_θ is a common form of encoding found in the visual cortex (44). This synergistic coding scheme has been shown to improve the representation of a stimulus, notably between motion direction and speed in area MT (45). Given the advantage conferred over independent tuning, it is not unexpected that multidimensional tuning would be exploited to form the substrate of precision weighting on sensory variables. Although the temporal nature of this code has been explored through temporal generalization (28), an intrinsic limit to our investigation is the lack of large-scale simultaneous recordings. Previous publications have reported population decoding by merging neural activity across electrodes or experiments (46, 47), which we validated in our data (Supplementary Figure 6).

Naturalistic stimuli have been used here to shed light on the processing of orientation precision. As the influence of orientation bandwidth on V1 is now more apprehensible in an instantaneous snapshot of visual inputs, the problem should be transposed to natural images and their exploration with saccadic behaviours, which adds a temporal dimension to the orientation distributions (48). Still, the functional organization of precision processing in V1 remains an open question. At the level of the orientation column, simultaneous layered recordings could pinpoint the origin of the various precision signals, which might also help understand the role that layer-dependent tuning diversity (49) plays in producing a resilient population code. On a larger scale, an interesting point is to be made in grasping how the co-dependency of orientation and precision codes are organized in functional maps. Even more so, species in which orientation selectivity is not map-based (50) could implement precision processing differently.

ACKNOWLEDGEMENTS

This article is dedicated to the memory of Umit Keysan (1992-2019), a dear colleague and wonderful friend. The authors would like to thank Genevieve Cyr for her technical assistance, Bruno Oliveira Ferreira de Souza and Visou Ady for advices regarding experimental procedures as well as Louis Eparvier and Jean-Nicolas Jérémie for their comments on the manuscript. This work was supported by ANR project "Horizontal-V1" ANR-17-CE37-0006 and by a CIHR grant to C.C (PJT-148959). H.J.L. was funded by an Ecole Doctorale 62 PhD grant.

DATA AVAILABILITY

The data generated in the present study are available from the corresponding author, H.J.L., upon reasonable request. No publicly available data was used in this study.

CODE AVAILABILITY

Data was analyzed using custom Python code, available at https://github.com/hugoladret, using the following libraries: SciPy (51), scikit-learn (52), numpy (53), PyTorch (54), lmfit (55) and Matplotlib (56).

AUTHOR CONTRIBUTIONS

L.U.P., C.C. F.C. and H.J.L. designed the study. H.J.L., N.C. and L.K. collected the data. H.J.L. and L.U.P. analyzed the data. H.J.L. and L.U.P. wrote the original draft of the manuscript. All authors reviewed and edited the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

References

1. Hubel, D. H. & Wiesel, T. N. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology* **148**, 574 (1959).
2. Priebe, N. J. Mechanisms of orientation selectivity in the primary visual cortex. *Annual review of vision science* **2**, 85–107 (2016).
3. Fiser, J., Chiu, C. & Weliky, M. Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature* **431**, 573–578 (2004).
4. Yoshida, T. & Ohki, K. Natural images are reliably represented by sparse and variable populations of neurons in visual cortex. *Nature communications* **11**, 1–19 (2020).
5. Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annual review of neuroscience* **24**, 1193–1216 (2001).
6. Field, D. J. Relations between the statistics of natural images and the response properties of cortical cells. *Josa a* **4**, 2379–2394 (1987).
7. Olshausen, B. A. & Field, D. J. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research* **37**, 3311–3325 (1997).
8. Froudarakis, E. *et al.* Population code in mouse v1 facilitates readout of natural scenes through increased sparseness. *Nature neuroscience* **17**, 851–857 (2014).
9. Olshausen, B. A. & Field, D. J. How close are we to understanding v1? *Neural computation* **17**, 1665–1699 (2005).
10. Felsen, G., Touryan, J., Han, F. & Dan, Y. Cortical sensitivity to visual features in natural scenes. *PLoS Biol* **3**, e342 (2005).
11. Rust, N. C. & Movshon, J. A. In praise of artifice. *Nature neuroscience* **8**, 1647–1650 (2005).
12. Portilla, J. & Simoncelli, E. P. A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision* **40**, 49–70 (2000).
13. Simoncini, C., Perrinet, L. U., Montagnini, A., Mamassian, P. & Masson, G. S. More is not always better: adaptive gain control explains dissociation between perception and action. *Nature neuroscience* **15**, 1596–1603 (2012).
14. Goris, R. L., Simoncelli, E. P. & Movshon, J. A. Origin and function of tuning diversity in macaque visual cortex. *Neuron* **88**, 819–831 (2015).
15. Ravello, C. R., Perrinet, L. U., Escobar, M.-J. & Palacios, A. G. Speed-selectivity in retinal ganglion cells is sharpened by broad spatial frequency, naturalistic stimuli. *Scientific reports* **9**, 1–16 (2019).
16. Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I. & Freeman, R. D. The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior. *Journal of neurophysiology* **57**, 773–786 (1987).
17. Baudot, P. *et al.* Animation of natural scene by virtual eye-movements evokes high precision and low noise in v1 neurons. *Frontiers in neural circuits* **7**, 206 (2013).
18. Julesz, B. Textons, the elements of texture perception, and their interactions. *Nature* **290**, 91–97 (1981).
19. Phillips, G. C. & Wilson, H. R. Orientation bandwidths of spatial mechanisms measured by masking. *JOSA A* **1**, 226–232 (1984).
20. Heeley, D., Timney, B., Paterson, I. & Thompson, R. Width discrimination for band-pass stimuli. *Vision research* **29**, 901–905 (1989).
21. Heeley, D. W. & Buchanan-Smith, H. M. The influence of stimulus shape on orientation acuity. *Experimental Brain Research* **120**, 217–222 (1998).
22. Keeble, D., Kingdom, F., Moulden, B. & Morgan, M. Detection of orientationally multimodal textures. *Vision Research* **35**, 1991–2005 (1995).
23. Beaudot, W. H. & Mullen, K. T. Orientation discrimination in human vision: Psychophysics and modeling. *Vision research* **46**, 26–46 (2006).
24. Berens, P. *et al.* A fast and simple population code for orientation in primate v1. *Journal of Neuroscience* **32**, 10618–10626 (2012).
25. Friston, K. & Kiebel, S. Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**, 1211–1221 (2009).

26. Leon, P. S., Vanzetta, I., Masson, G. S. & Perrinet, L. U. Motion clouds: model-based stimulus synthesis of natural-like random textures for the study of motion perception. *Journal of neurophysiology* **107**, 3217–3226 (2012).
27. Naka, K. & Rushton, W. A. S-potentials from colour units in the retina of fish (cyprinidae). *The Journal of physiology* **185**, 536–555 (1966).
28. King, J.-R. & Dehaene, S. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in cognitive sciences* **18**, 203–210 (2014).
29. Zhang, K. & Sejnowski, T. J. Neuronal tuning: To sharpen or broaden? *Neural computation* **11**, 75–84 (1999).
30. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).
31. Karklin, Y. & Lewicki, M. S. Emergence of complex cell properties by learning to generalize in natural scenes. *Nature* **457**, 83–86 (2009).
32. Schwartz, O., Pillow, J. W., Rust, N. C. & Simoncelli, E. P. Spike-triggered neural characterization. *Journal of vision* **6**, 13–13 (2006).
33. Chance, F. S., Nelson, S. B. & Abbott, L. F. Complex cells as cortically amplified simple cells. *Nature neuroscience* **2**, 277–282 (1999).
34. Benvenuti, G. *et al.* Scale-invariant visual capabilities explained by topographic representations of luminance and texture in primate v1. *Neuron* **100**, 1504–1512 (2018).
35. Bringuier, V., Chavane, F., Glaeser, L. & Frégnac, Y. Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science* **283**, 695–699 (1999).
36. Carandini, M. & Ringach, D. L. Predictions of a recurrent model of orientation selectivity. *Vision research* **37**, 3061–3071 (1997).
37. Monier, C., Chavane, F., Baudot, P., Graham, L. J. & Frégnac, Y. Orientation and direction selectivity of synaptic inputs in visual cortical neurons: a diversity of combinations produces spike tuning. *Neuron* **37**, 663–680 (2003).
38. Rao, R. P. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience* **2**, 79–87 (1999).
39. Wang, C., Huang, J. Y., Bardy, C., FitzGibbon, T. & Dreher, B. Influence of 'feedback'signals on spatial integration in receptive fields of cat area 17 neurons. *Brain research* **1328**, 34–48 (2010).
40. Ito, M. & Gilbert, C. D. Attention modulates contextual influences in the primary visual cortex of alert monkeys. *Neuron* **22**, 593–604 (1999).
41. Kanai, R., Komura, Y., Shipp, S. & Friston, K. Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, 20140169 (2015).
42. Liang, H. *et al.* Interactions between feedback and lateral connections in the primary visual cortex. *Proceedings of the National Academy of Sciences* **114**, 8637–8642 (2017).
43. de Souza, B. O. F., Cortes, N. & Casanova, C. Pulvinar modulates contrast responses in the visual cortex as a function of cortical hierarchy. *Cerebral Cortex* **30**, 1068–1086 (2020).
44. Zohary, E. Population coding of visual stimuli by cortical neurons tuned to more than one dimension. *Biological cybernetics* **66**, 265–272 (1992).
45. Macellaio, M. V., Liu, B., Beck, J. M. & Osborne, L. C. Why sensory neurons are tuned to multiple stimulus features. *bioRxiv* (2020).
46. Quiroga, R. Q., Reddy, L., Koch, C. & Fried, I. Decoding visual inputs from multiple neurons in the human temporal lobe. *Journal of neurophysiology* **98**, 1997–2007 (2007).
47. Guitchounts, G., Masis, J., Wolff, S. B. & Cox, D. Encoding of 3d head orienting movements in the primary visual cortex. *Neuron* **108**, 512–525 (2020).
48. Friston, K., Adams, R., Perrinet, L. & Breakspear, M. Perceptions as hypotheses: saccades as experiments. *Frontiers in psychology* **3**, 151 (2012).
49. Ringach, D. L., Shapley, R. M. & Hawken, M. J. Orientation selectivity in macaque v1: diversity and laminar dependence. *Journal of Neuroscience* **22**, 5639–5651 (2002).
50. Hansel, D. & van Vreeswijk, C. The mechanism of orientation selectivity in primary visual cortex without a functional map. *Journal of Neuroscience* **32**, 4049–4064 (2012).
51. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272 (2020).
52. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
53. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272 (2020).
54. Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. In Wallach, H. *et al.* (eds.) *Advances in Neural Information Processing Systems 32*, 8024–8035 (Curran Associates, Inc., 2019).
55. Newville, M. *et al.* Lmfit: Non-linear least-square minimization and curve-fitting for python. *ascl* ascl–1606 (2016).
56. Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering* **9**, 90–95 (2007).
57. Villeneuve, M. Y. & Casanova, C. On the use of isoflurane versus halothane in the study of visual response properties of single cells in the primary visual cortex. *Journal of neuroscience methods* **129**, 19–31 (2003).
58. Siegle, J. H. *et al.* Open ephys: an open-source, plugin-based platform for multichannel electrophysiology. *Journal of neural engineering* **14**, 045003 (2017).
59. Pachitariu, M., Steinmetz, N. A., Kadir, S. N., Carandini, M. & Harris, K. D. Fast and accurate spike sorting of high-channel count probes with kilosort. In *Advances in neural information processing systems*, 4448–4456 (2016).
60. Rossant, C. *et al.* Spike sorting for large, dense electrode arrays. *Nature neuroscience* **19**, 634–641 (2016).
61. Peirce, J. *et al.* Psychopy2: Experiments in behavior made easy. *Behavior research methods* **51**, 195–203 (2019).
62. Carandini, M. & Heeger, D. J. Normalization as a canonical neural computation. *Nature Reviews Neuroscience* **13**, 51–62 (2012).
63. Movshon, J. A., Thompson, I. & Tolhurst, D. Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *The Journal of physiology* **283**, 101–120 (1978).
64. Smith, M. A., Majaj, N. J. & Movshon, J. A. Dynamics of motion signaling by neurons in macaque area mt. *Nature neuroscience* **8**, 220–228 (2005).
65. Swindale, N. V. Orientation tuning curves: empirical description and estimation of parameters. *Biological cybernetics* **78**, 45–56 (1998).
66. Bishop, C. M. *Pattern recognition and machine learning* (springer, 2006).
67. Brodersen, K. H., Ong, C. S., Stephan, K. E. & Buhmann, J. M. The balanced accuracy and its posterior distribution. In *2010 20th International Conference on Pattern Recognition*, 3121–3124 (IEEE, 2010).

Materials and Methods

Animal procedures. Experiments were performed on 3 adult cats (3.6 - 6.0 kg, 2 males). All surgical and experimental procedures were carried out in compliance with the guidelines of the Canadian Council on Animal Care and were approved by the Ethics Committee of the University of Montreal (CDEA #20-006). Animals were first administered atropine (0.1 mg/kg) and acepromazine (Atravet, 1mg/kg) subcutaneously to reduce the parasympathetic effects of anesthesia and provoke sedation, respectively. Anesthesia was induced with 3.5% Isoflurane in a 50:50 (v/v) mixture of O_2 and N_2O. Isoflurane concentration was maintained at 1.5% during surgical procedures. A tracheotomy was performed and animals were immobilized using an intravenous injection of 2% gallamine triethiodide. Animals were then artificially ventilated and a 1:1 (v/v) solution of 2% gallamine triethiodide (10 mg/kg/h) in 5% of dextrose in lactated ringer solution was continuously administered to maintain muscle relaxation. Throughout the experiment, expired level of CO2 was maintained between 35 and 40 mmHg by adjustement of the tidal volume and respiratory rate. Heart rate was monitored and body temperature was maintained at 37°C using a feedback controlled heated blanket. Dexamethasone (4mg) was administered intramuscularly every 12h to reduce cortical swelling. Pupils were dilated using atropine (Mydriacyl) and nictitaing membranes were retracted using phenylephrine (Midfrin). Rigid contact lenses of appropriate power were used to correct the eyes' refraction and eye lubricant was used to avoid corneal dehydratation. Lidocaine hydrochloride (2%) was used in all incisions and pressure points. A craniotomy was performed between Horsley-Clarke coordinates 4-8 P ; 0.5 - 2 L to access the area 17 (V1) contralateral to the stimulation side. Small durectomies were performed for each electrode penetration. A 2% agar solution in saline was applied over the exposed regions to stabilize recordings and avoid the drying of cortical surface.

Electrophysiological recordings. During recording sessions, anesthesia was changed to 0.5-1% Halothane, as isoflurane has been shown to yield a depression of visual responses (57). Extracellular activity was recorded using 32 channel linear probes (≈1 MΩ, 1×32-6 mm-100-177, Neuronexus) and acquired at 30 KHz using an Open Ephys acquisition board (58). Single units were isolated using Kilosort spike sorting software (59) and manually curated using the phy software (60). Clusters of low amplitude templates or ill-defined margins were excluded from analysis. Furthermore, clusters were excluded if their firing rate dropped below 5 spikes.s^{-1} for more than 30 seconds or if their tuning curve was poorly fitted with a Von-Mises distribution ($r^2 < .75$), leaving 254 putative neurons for further analysis.

Visual Stimulation. Visual stimuli were generated using Psychopy (61) and were projected monocularly with a PROPixx projector (VPixx Technologies Inc., St-Bruno, QC, Canada) onto an isoluminant screen (Da-Lite© screen) located 57 cm from the animal's eye, covering 104°x 79° of visual angle with a mean luminance of 25 cd/m^2. The stimuli used here, Motion Clouds (26), are a class of band-pass filtered white-noise textures (62) that provide parametric control over the content of the stimuli while retaining the statistics of natural images (6). The envelope of the filters in Fourier space is a Gaussian in the coordinates of the relevant axis, in which it is described by its mean and bandwidth. As such, a MotionCloud (MC) is defined as:

$$MC = \mathcal{F}^{-1}(G(f_0, B_f) \cdot O(\theta, B_\theta)) \quad \textbf{(1)}$$

where $\mathcal{F}$ is the Fourier transform, G the spatial frequency envelope and O the orientation envelope. The spatial frequency envelope follows a log-normal distribution:

$$G(f_0, B_f)(f_x, f_y) = \frac{1}{f_r} \exp\left\{ -\frac{1}{2} \left(\frac{\ln(\frac{f_r}{f_0})}{\ln(\frac{f_0+B_f}{f_0})} \right)^2 \right\} \quad \textbf{(2)}$$

where f_0 is the mean spatial frequency and B_f is the bandwidth of the spatial frequency distribution, both in cycles per degree. The orientation envelope is a Von-Mises distribution:

$$O(\theta, B_\theta)(f_x, f_y) = \exp\left\{ \frac{\cos(2(\theta_f - \theta))}{4 \cdot B_\theta^2} \right\} \quad \textbf{(3)}$$

where θ_f is the angle of f_x, f_y in the Fourier plane, θ is the mean orientation and B_θ is the bandwidth of the orientation distribution. For narrow distributions, this distribution is close to a Gaussian and B_θ measures its standard deviation. The spatial frequency parameters were set at $f_0 = B_f = 0.9$ cpd and orthogonal drift speed was set to 10°/s, within the response range of area 17 neurons (63). For the orientation envelope, θ was varied in 12 even steps from 0 to π rad and B_θ in 8 even steps from $\frac{\pi}{5}$ to ≈ 0. All stimuli were displayed at 100% contrast. Each Motion Cloud was displayed for 300 ms, interleaved with the presentation of a mean luminance screen for 150 ms. Trials were fully randomized and each stimulus was presented 15 times.

Single Neuron Analysis. Tuning curves were generated by selecting a 300 ms window maximizing spike-count variance (64) in which firing rate was averaged and baseline substracted. Tuning curves were averaged across drift directions. A Poisson loss function was minimized to fit a Von-Mises distribution to the data:

$$f(\theta_k) = R_0 + (R_{\max} - R_0) \cdot \exp\{\kappa \cdot (\cos(2(\theta_k - \theta)) - 1)\} \quad \textbf{(4)}$$

where $R_{\max}$ is the response at the preferred orientation θ, R_0 the response at the orthogonal orientation, κ is a measure of concentration and θ_k the orientation of the stimuli. A local measure of orientation tuning around the peak of the function, the half-width at half height (HWHH) (65) was measured as:

$$\text{HWHH} = \frac{1}{2} \arccos\left(1 + \frac{1}{\kappa} \ln\left(\frac{1 + \exp\{-2 \cdot \kappa\}}{2}\right)\right) \quad \textbf{(5)}$$

A global measure of the orientation tuning was also assessed by computing the Circular Variance (CV) of the raw tuning curve:

$$\text{CV} = 1 - \left| \frac{\sum_k R(\theta_k) \exp\{2i\theta_k\}}{\sum_k R(\theta_k)} \right| \quad \textbf{(6)}$$

where $R(\theta_k)$ is the response to a stimuli of angle θ_k. CV varies from 1 for exceptionally orientation-selective neurons to 0 for untuned neurons (49).

The evolution of HWHH and CV as functions of B_θ were assessed with a Naka-Ruhston function:

$$f(B_\theta) = f_0 + f_{\max} \cdot \frac{B_\theta^n}{B_\theta^n + B_{\theta 50}^n} \quad \textbf{(7)}$$

where f_0 is the base value, $f_{\max}$ the maximal value, $B_{\theta 50}$ the orientation bandwidth at half $f_{\max}$ and n a strictly positive exponent of the function (27). A Wilcoxon signed-rank test, corrected for continuity, was used to assess the significance of orientation selectivity. Significance of the orientation tuning was computed by comparing the firing rate at the preferred and orthogonal orientations for all trials. Changes of preferred orientation were assessed as the difference between $\arg\max_{B_\theta = 0.0°}$ and $\arg\max_{B_\theta \text{ untuned}}$, where $B_{\theta \text{ untuned}}$ is the broadest orientation bandwidth to which a neuron remains significantly orientation tuned, following the previous test. Variations of peak amplitude was measured by comparing the firing rate at preferred orientation for $B_\theta = 0.0°$ and $B_\theta = B_{\theta_{\text{untuned}}}$.

Population Decoding. The stimulus identity was decoded using a multinomial logistic regression classifier (66), which can be interpreted as an instantaneous, neuronal plausible readout of the population activity (24). For a given stimuli, population activity was a vector $X(t) = \begin{bmatrix} X_1(t) & X_2(t) & \cdots & X_{254}(t) \end{bmatrix}$, where $X_i(t)$ is the spike count of neuron i in time window $[t; t + \Delta T]$, where t is the time bin and ΔT the size of the integration window, usually 100 ms. Time t was varied from -200 ms to 400 ms in steps of 10 ms, relative to the stimulation onset. Each time bin was labeled as the end of the time window and was hence reported as $t + \Delta T$.

The multinomial logistic regression is an extension of the binary logistic regression (66) which was trained to classify the neural activity vector between K possible classes. The probability of such vector to belong to a given class is:

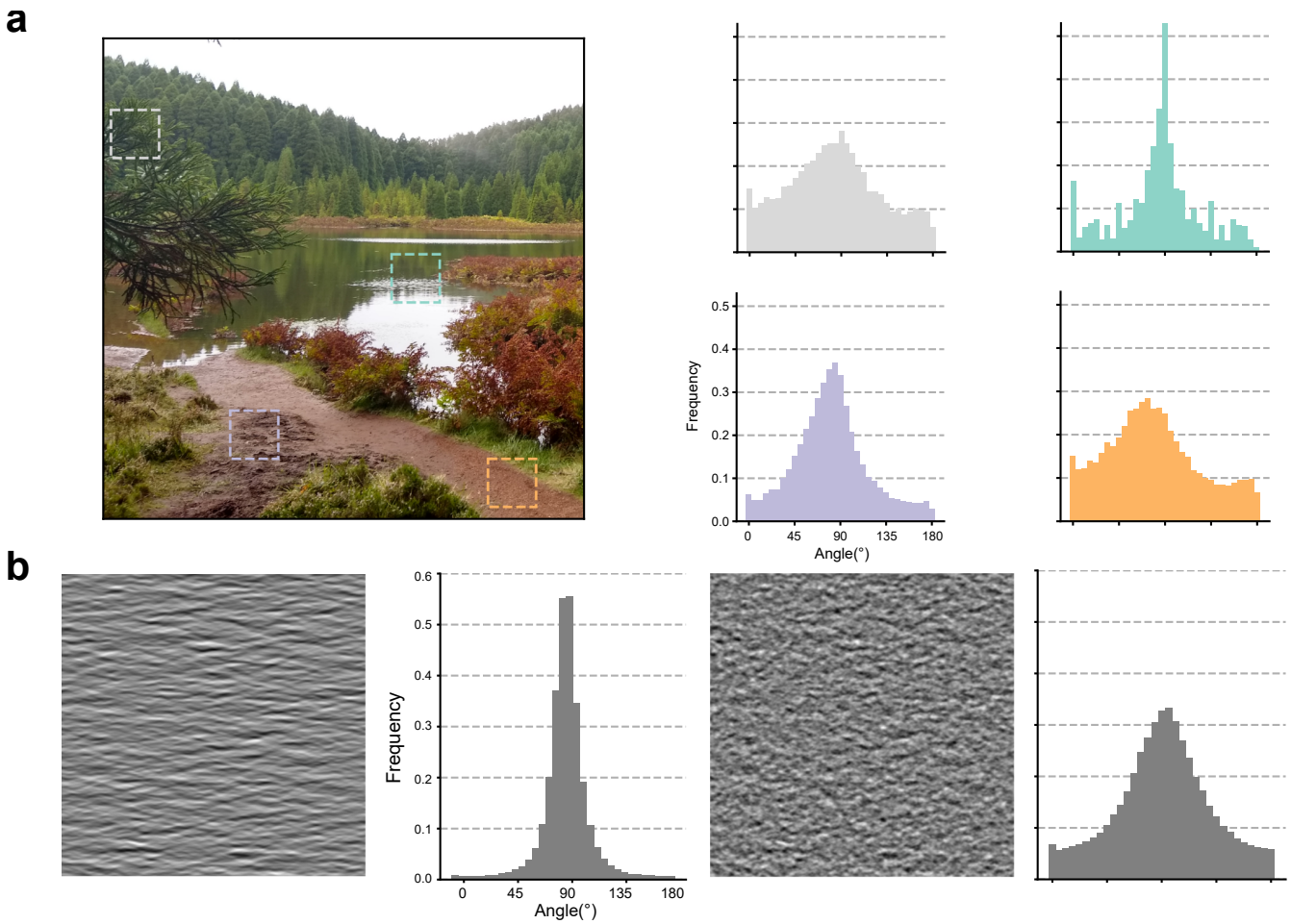
$$P(y = k | X(t)) = \frac{\exp\{\langle \beta_k, X(t) \rangle\}}{\sum_{k'=1}^{K} \exp\{\langle \beta_{k'}, X(t) \rangle\}} \quad \textbf{(8)}$$

where $\langle \cdot, \cdot \rangle$ is the scalar product over the different neurons, $k = 1, \ldots, K$ is the class, out of K possible values and β_k are the vectors of learned coefficients of the classifier. We trained several such classifiers, to decode orientation θ ($K = 12$), orientation bandwidth B_θ ($K = 8$) or both ($K = 12 \times 8 = 96$). All meta-parameters (the time window size, penalty type, regularization strength and train/test split size) were controlled to show that the decoder is optimally parameterized and that its results are dependent on experimental data, not on decoder parameterization (Supplementary Figure 5).

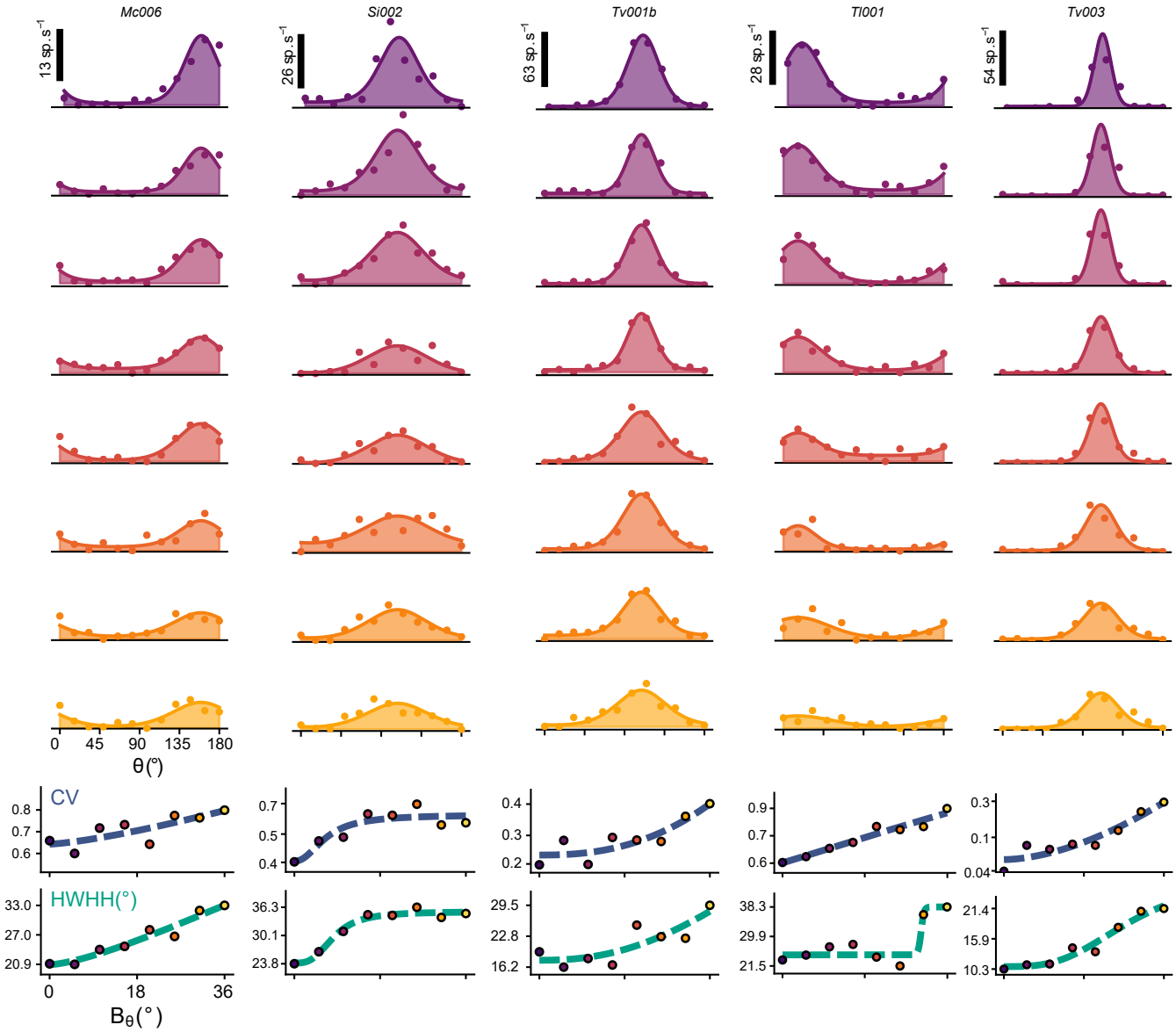
Decoding accuracy was reported as the average accuracy across classes, also known as the balanced accuracy score (67), which accounts for possible imbalances in the learning or testing dataset. Decoding performance was also reported with confusion matrices, in which the values on each row i and column j represents the normalized number of times a stimuli of class $k = i$ is predicted to belong to the class $k = j$. Hence, a perfect decoder would produce a perfectly diagonal confusion matrix (unit matrix). When reporting a confusion matrix, color maps values were clipped between chance level and maximum accuracy. The temporal evolution of the balanced accuracy score corresponds to the mean of the diagonal of the confusion matrix at each time bin. We reported the significativity of decoders by splitting the population activity in separate training and testing data sets, then performing 6 different such splits and comparing the resulting confusion matrices to chance level decoding matrices.

Temporal Generalization. Temporal generalization assessed the dynamics of the neural code by training and testing the decoder at different time bins (28). The resulting matrix displayed the capacity of a decoder to temporally generalize a structure it learned from the data across two time points. The diagonal, where $t_{\text{train}} = t_{\text{test}}$, was the normal time course of the decoder's accuracy. The upper half of the matrix corresponds to transposition points where the code is tested at a prior time step compared to its learning step, and conversely for the lower half. Hence, an asymmetric matrix postulates a neural code that can be either transient, transitioning between representations (lower half > upper half, large contoured regions) or stable through the trial (upper half > lower half, sparse contoured regions). Where significativity is reported, 6 different train/test splits were performed for each transposition, and the significance was computed between the difference of symmetrical points around the time identity line using a permutation test (1000 permutations).

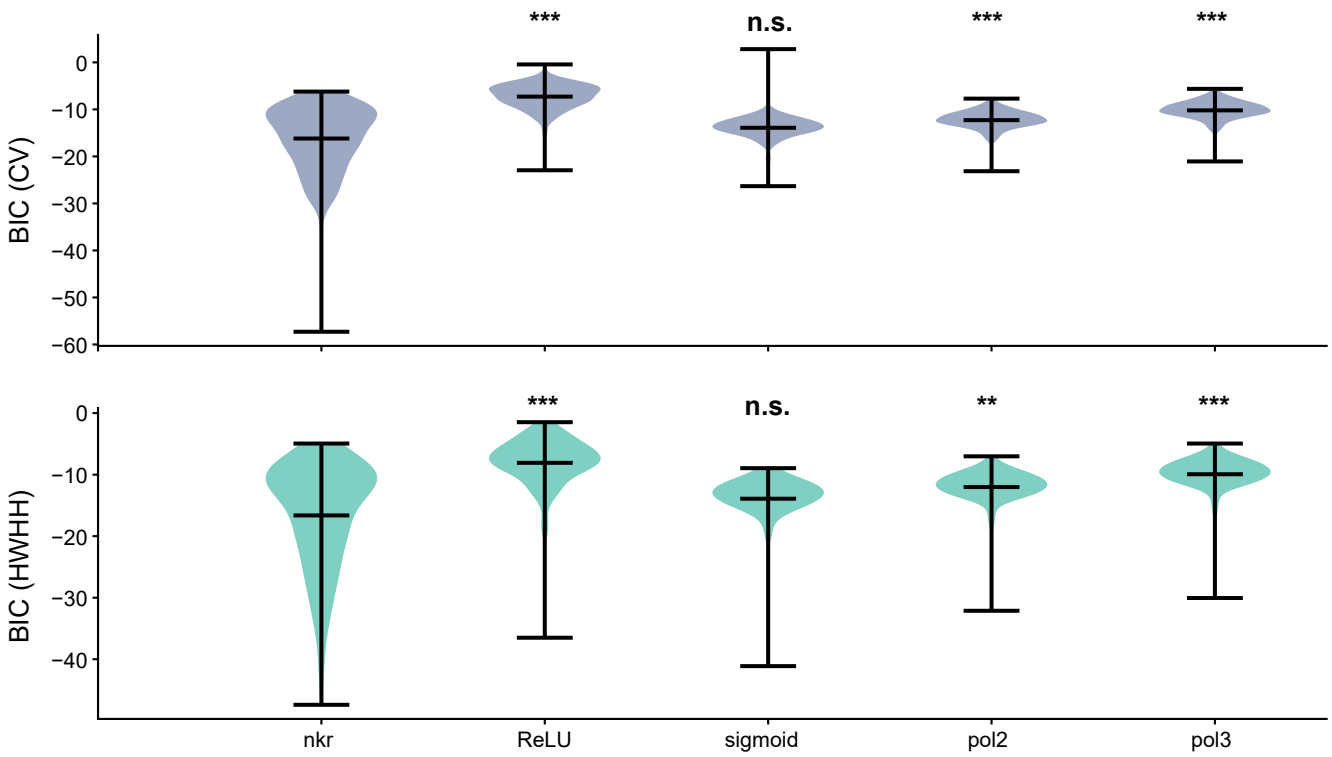
Supplementary Figures



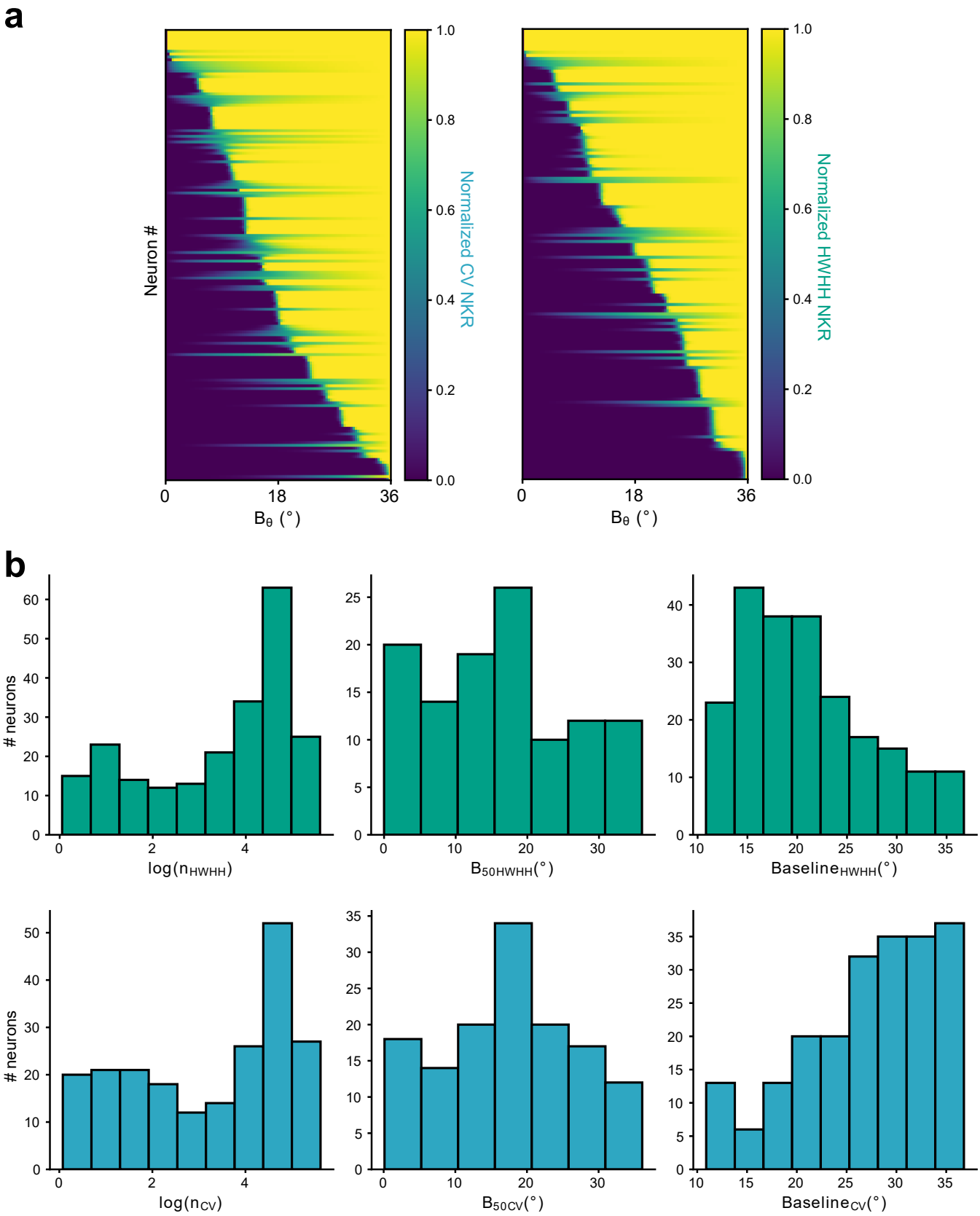
Supplementary Figure 1. Motion Clouds capture the variety of orientation distributions found in natural images. (**a**) Distribution of orientations found in a natural image (São Miguel Island, Azores, Portugal, picture taken by H.J.L.), retrieved with a Histogram of Oriented Gradients (16x16 cell size, 200^2 pixel subset of the image). (**b**) Two Motion Clouds, oriented at 90° relative to the horizontal axis, with orientation bandwidths of respectively $B_\theta = 15.4°$ and $B_\theta = 36.0°$.



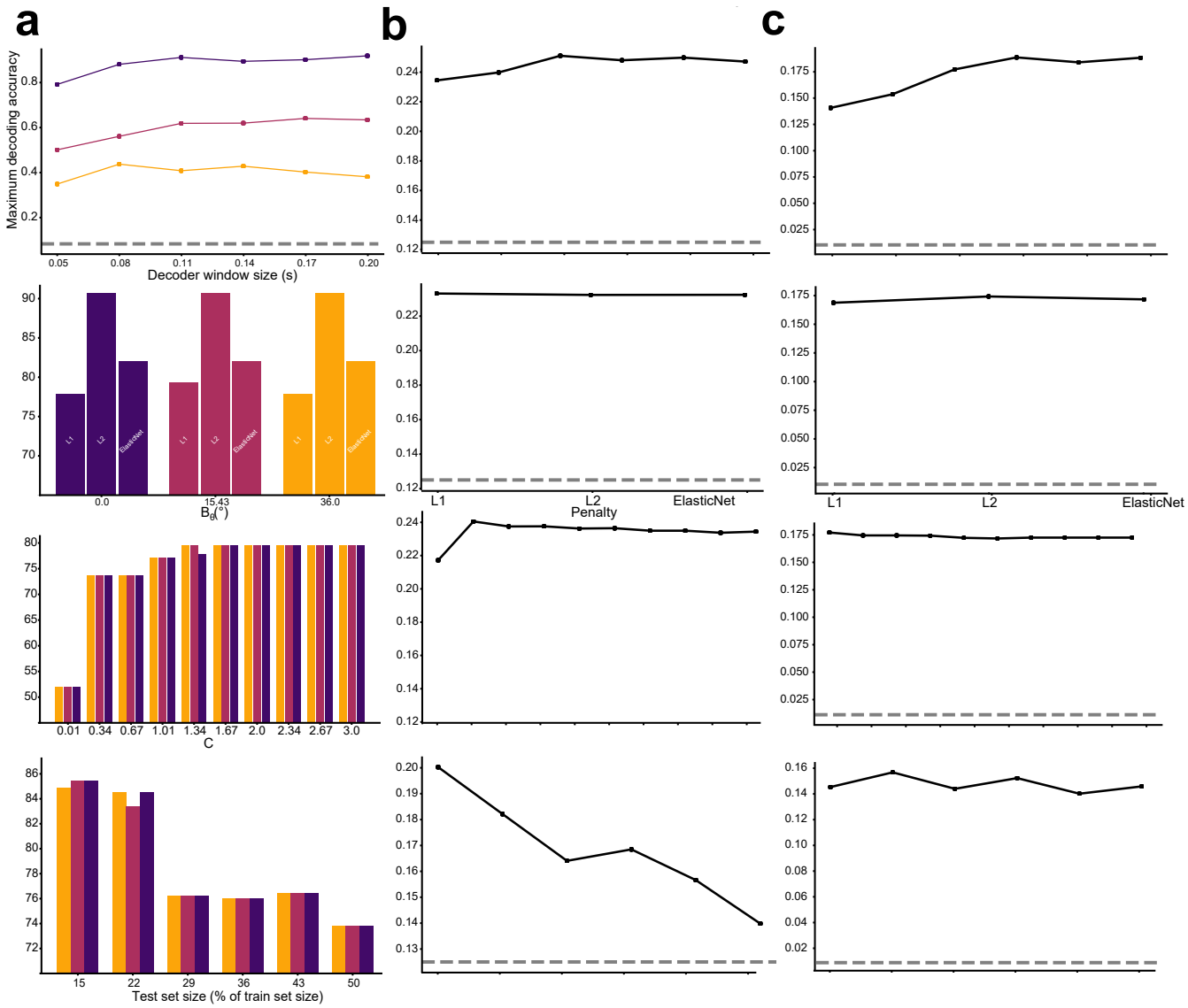
Supplementary Figure 2. Additional examples of single neuron responses. Top: Tuning curves of 5 additional neurons in response to Motion Clouds of decreasing B_θ (top to bottom). Bottom: Changes in Half-Width at Half-Height (HWHH) and Circular Variance (CV) as a function of B_θ for the respective neurons above and fitted with Naka-Rushton functions (dashed curves).



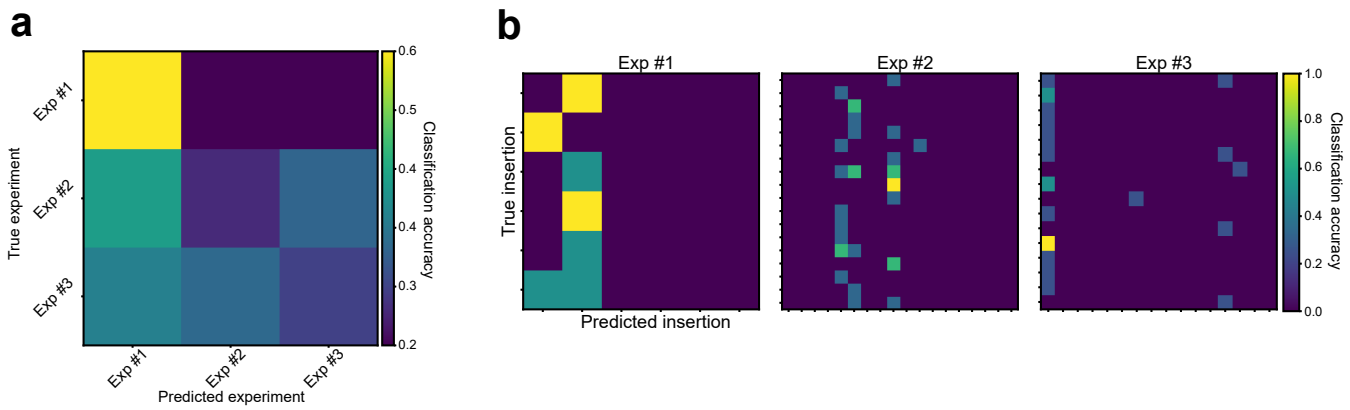
Supplementary Figure 3. Criterion on the model selection for HWHH and CV curves. Top: Violin plot of the Bayesian Information Criterion (BIC) of the CV curves of all recorded neurons. Each violin plot represent a different kind of fitted equation, respectively: Naka-Rushton (see Materials and Methods) ; Rectified Linear Unit (ReLU, $f(x) = max(0, x)$) ; sigmoid ($f(x) = \frac{e^x}{e^x+1}$) ; second degree polynomial function ($f(x) = ax^2 + bx + c$) and third degree polynomial function ($f(x) = ax^3 + bx^2 + cx + d$). A lower BIC indicates less information lost in the fitting process, hence a better fitting model. Bottom: Violin plot of the BIC for the HWHH curves. n.s., not significant, * p < 0.05, ** p < 0.01, *** p < 0.001 (Kruskal-Wallis H-test, post-hoc Dunn Pairwise test, Bonferroni corrected)



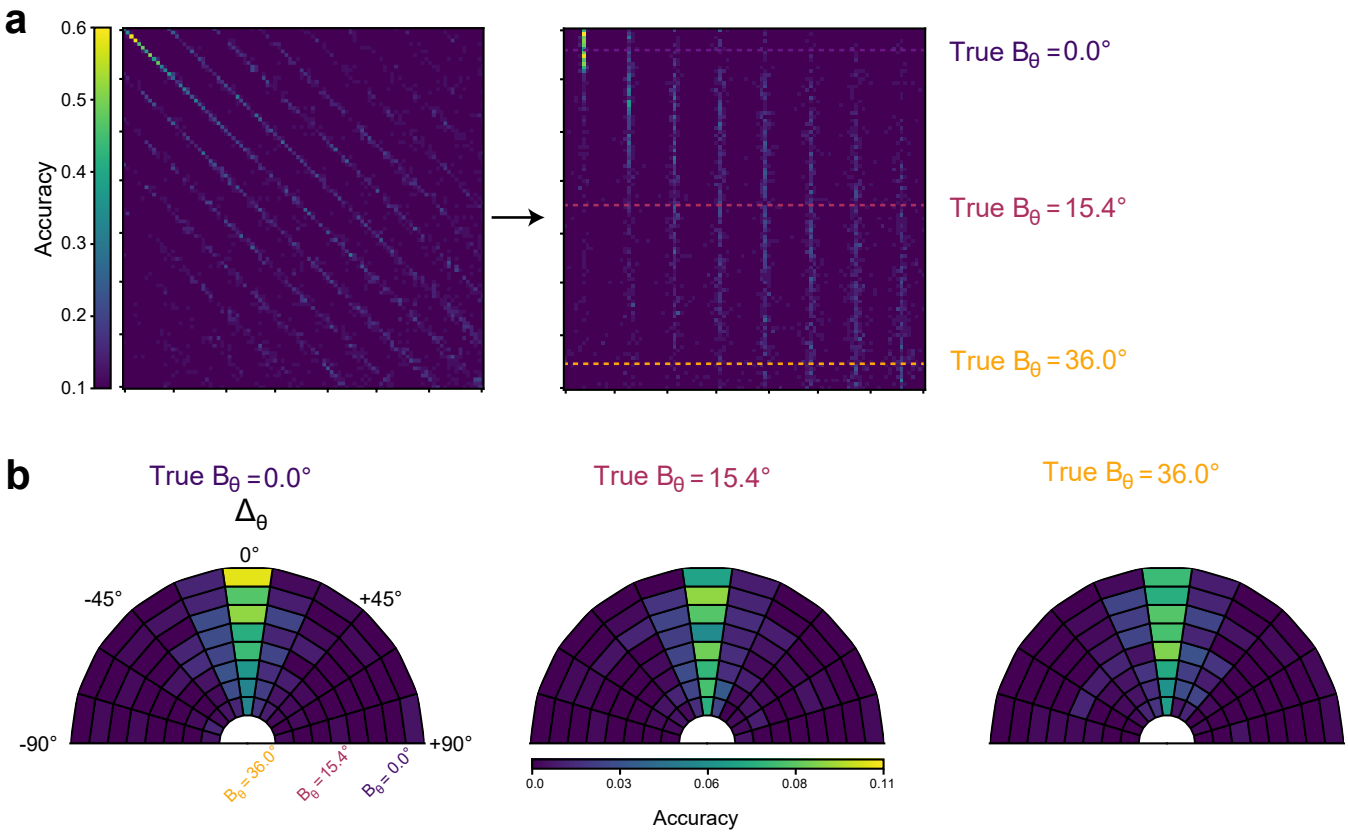
Supplementary Figure 4. Population distribution of the Naka-Rushton (NKR) fits. (**a**) Representation of the NKR of the population as a colored mesh. Each row corresponds to the normalized fitted function of a neuron. Neurons whose B_{50} were greater than $36°$ are not shown, for clarity's sake (65 neurons hidden for HWHH, 49 for CV). (**b**) Histograms of the population distributions of NKR parameters. Top: Histograms of the parameters for the HWHH values, from left to right : log of n, the exponent parameter, B_50, the change point parameter and the baseline, the value of the NKR for $B_\theta = 0.0°$. Bottom : Histograms of the parameters for the CV values.



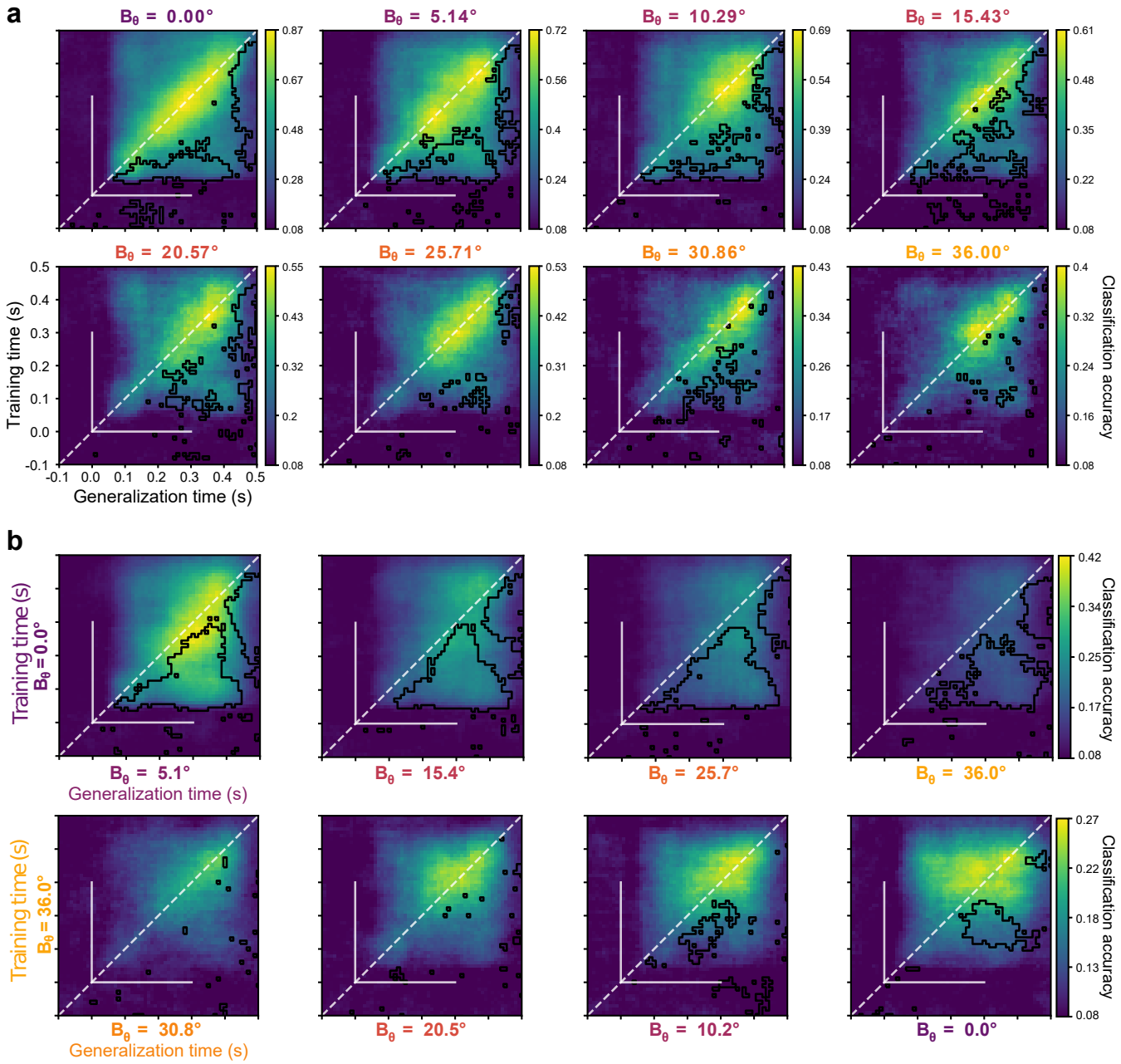
Supplementary Figure 5. Parameters evaluation of the decoders. (**a**) Parameter scanning of the orientation θ decoders for three B_θ. Maximum accuracy reached during the time course of each decoder was reported for each parameter. From top to bottom, the parameters optimized are: the length of the time window ΔT, the type of penalization norm applied to the decoder, the regularization strength parameter and the percentage of data kept out of the training set to evaluate the decoder's accuracy. (**b**) Parameters of the orientation bandwidth B_θ decoder. (**c**) Parameters of the orientation and orientation bandwidth $B_\theta \times \theta$ decoder.



Supplementary Figure 6. Decoding the experiment identity and insertion identity of neurons. (**a**) Confusion matrix of a decoder trained to retrieve the experiment identity using three groups of 30 neurons (bootstrapped 1000 times). (**b**) Confusion matrices of three decoders trained to retrieve the insertion identity of the neurons recorded in each experiments.



Supplementary Figure 7. Example of polar histograms of a confusion matrix of the $B_\theta * \theta$ decoder. (**a**) Starting from the Figure 3b representation, each row is shifted by its θ value. A given row of the confusion matrix (actual B_θ values) is then marginalized over all the columns (predicted B_θ values), resulting in an 8 B_θ by 12 θ matrix. (**b**) Polar representation of the Figure 3b, $t = 400$ ms, confusion matrix after aforementioned marginalization. The leftmost histogram is the marginalization of the upper row of the matrix. This polar plot shows that population activity elicited by $B_\theta = 0°$ Motion Clouds correctly creates a prediction of stimuli of narrow B_θ. Similarly, the rightmost polar representation is the lower row of the confusion matrix and shows a decoding of Motion Clouds of broad B_θ. Δ_θ is a measure of the decoding's broadness in the orientation domain and represents the events where a tilted orientation was erroneously decoded compared to the actual stimulus orientation ($\Delta_\theta = 0°$).



Supplementary Figure 8. Temporal generalization of all orientation decoders. Contoured regions indicate significant difference (1000 permutation test, $n = 6$, threshold $p < 0.05$) compared to the symmetrical region with respect to the time identity line (dashed white). White lines represent stimulation duration. (**a**) Temporal generalization matrices of decoders trained and generalized on the same B_θ, whose difference between upper and lower halves around the time identity line are shown in Figure 5a. (**b**) Temporal generalization matrices of decoders trained and tested on different B_θ, with upper and lower rows corresponding to the uppper and lower rows of Figure 5b.