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 at all bandwidths. This decoder demonstrates that V1 population co-encodes orientation and its precision, which enhances population decoding performances. This internal representation is mediated by temporally distinct neural dynamics and supports a precision-weighted description of 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 understanding of how the early visual system works has to also account for the behavior 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). On the contrary, the cortical processing of the broadness of this distribution (i.e. the orientation bandwidth) is unclear. 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 on vision come from the field of texture perception (17). Intuitively, psychophysics experiments have shown that increased broadness in orientation distributions worsens orientation discrimination performances (18–20), which can be explained by computational models of intracortical dynamics (21–23). Despite these converging observations, there has been very few investigations in the neural correlates of the processing of orientation bandwidth. Recent work by Goris *et al.* (14) has shown that orientation bandwidth causes tuning modulations of single neurons in V1 and V2 that are coherent with previous psychophysics and computational studies. These modulations are heterogeneous at the population level and provide a robust theoretical basis for encoding natural images. However, the detailed processes involved in this population code remain unknown.

Here, we used naturalistic stimuli to study the encoding of the bandwidth 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 bandwidth is not encoded explicitly in the activity of individual V1 neurons, we showed that it is actually co-encoded with orientation, which improves the orientation encoding performances of a neural population. Moreover, the bandwidth of an orientation is linked to its processing time, which stems from temporally separated neural representations of orientation bandwidths within 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 of milliseconds later. These results suggest that the precision of the orientation, in the form of its bandwidth, is actively processed in V1. This supports 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

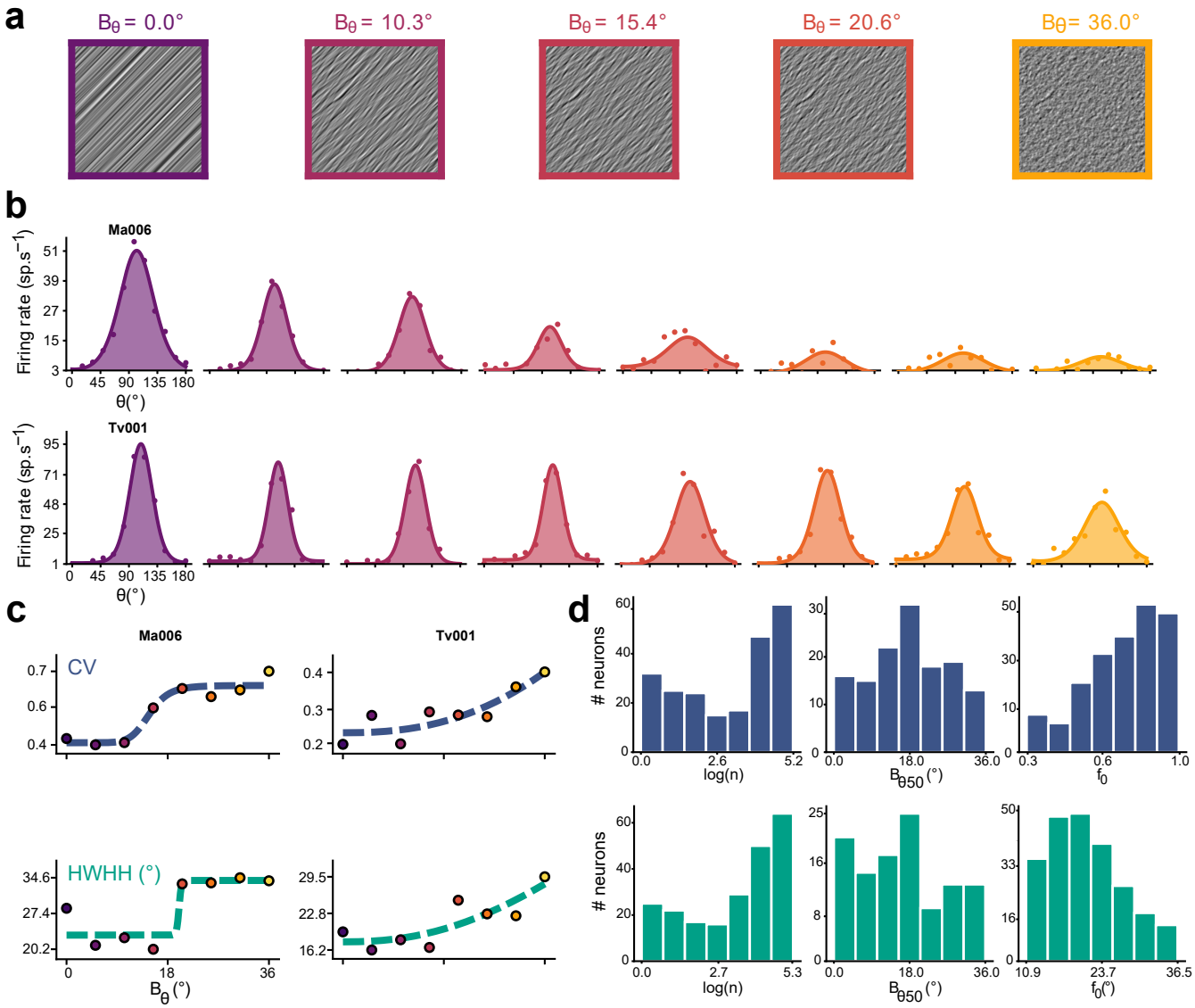


Figure 1. Band-pass filtered white noise stimuli and single-neuron responses. (**a**) Examples of Motion Clouds, oriented at 45° relative to the horizontal axis. Neuronal responses were measured with orthogonally drifting Motion Clouds of 12 different orientations θ and 8 orientation bandwidth B_θ. (**b**) Tuning curves of two examples neurons (labelled 'Ma006' and 'Tv001') stimulated with Motion Clouds of increasing B_θ, i.e. increasing orientation bandwidth (from left to right). Colored points represent the mean firing rate across trials (baseline subtracted, 300 ms average), with lines indicating a fitted von Mises function. (**c**) Precision-tuning functions (PTF), assessing the changes of orientation tuning as measured by the Half-Width at Half-Height (HWHH) and Circular Variance (CV) as a function of B_θ, fitted with Naka-Rushton (NKR) functions (dashed curves). (**d**) Distribution of the NKR parameters for the entire recorded population.

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 median orientation θ and the bandwidth of the orientation distribution B_θ (Figure 1a), which as the standard deviation of the distribution is related to the variance or inverse precision of the orientation input to V1. Like natural images, Motion Clouds are complex mixtures of orientations, but their stationarity in the spatial domain removes any potential 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). Specifically, neuron Ma006 was no longer tuned to stimuli of $B_\theta = 36.0°$ ($p = 0.24$, Wilcoxon signed-rank test, firing rate of preferred vs orthogonal orientation), contrary to neuron Tv001 which remained orientation selective

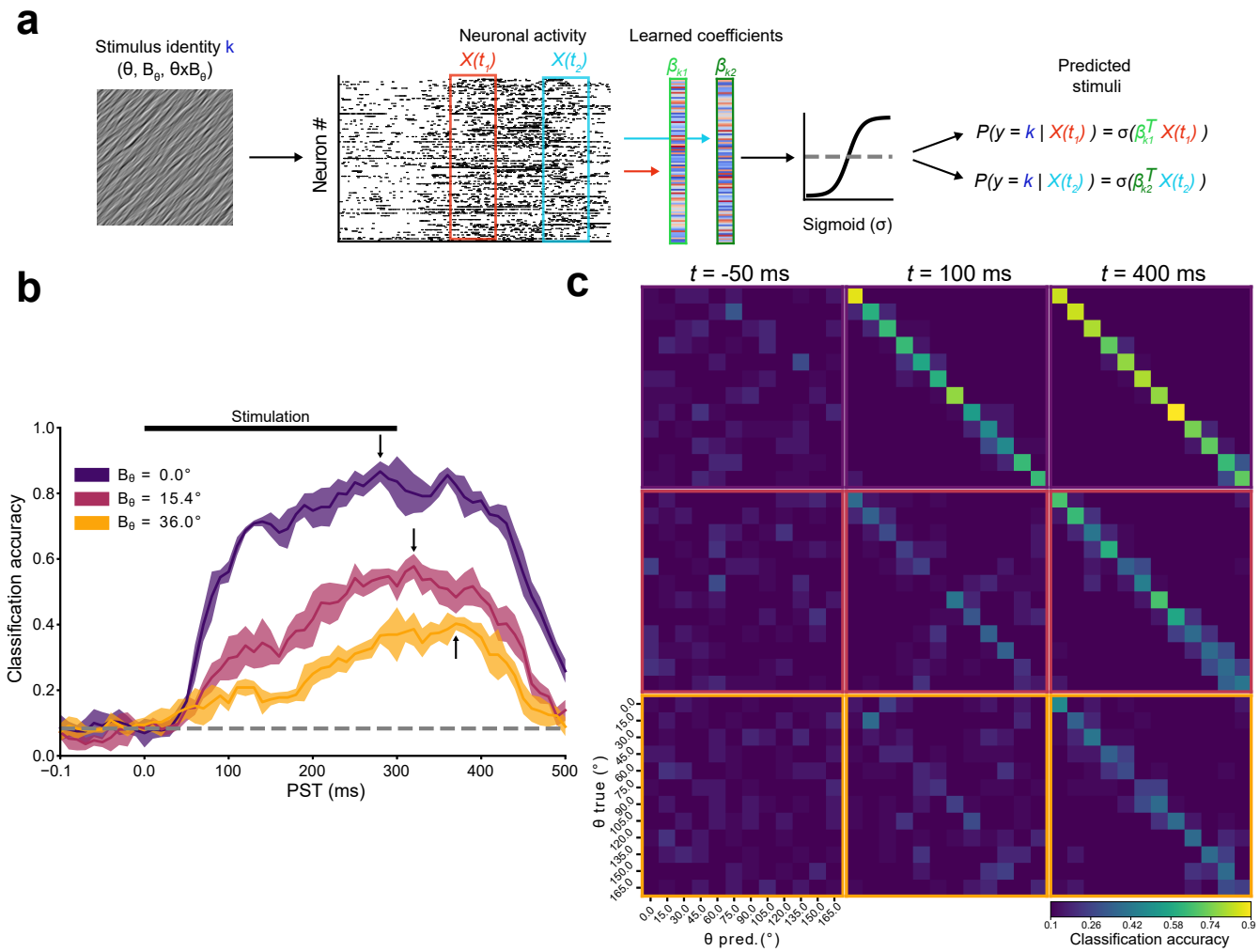


Figure 2. Decoding orientation from population activity. (**a**) Illustration of the decoding model. For each trial, spikes from all single units in a given time window were pooled in a single neural activity vector $X(t)$. A multinomial logistic classifier predicts the identity k of the stimulus in this trial by estimating the probability for different hypothesis y. This is achieved by learning a weight vector β_k which maximizes the accuracy of the prediction $P(y = k|X(t)) \approx \sigma(\beta_k^T X(t))$, where σ is the sigmoid function. The resulting accuracy of this classifier 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 on 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 is framed with the color code of the corresponding decoder in (b).

($p = 10^{-6}$). 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 various degrees of non-linear observed in the changes of tuning (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. First, the bandwidth value $B_{\theta 50}$ corresponds to the B_θ at which a neuron transitioned towards a more broadly tuned state. Secondly, the slope n, which indicates the steepness of the variations of tuning as B_θ increases. This variable was correlated to the tuning width measured at $B_\theta = 0.0°$ ($p_{\text{HWHH}} = 0.048$, Spearman's $R_{\text{HWHH}} = 0.12$, $p_{\text{CV}} = 10^{-6}$, Spearman's $R_{\text{CV}} = 0.33$), that is, broadly tuned neurons underwent tuning modulations at lower B_θ compared to narrower neurons, similar to previous experimental observations (14). No other significant relationship was found, except for an $n; B_{\theta 50}$ correlation produced by the Naka-Rushton fitting procedure (the optimal fit of a neuron whose $B_{\theta 50}$ is past the maximum tested B_θ is a linear function). Overall, the population's parameters were heterogeneous (Figure 1c), suggesting a diversity of tuning modulations between neurons would provide better orientation selectivity capacity between multiple orientation bandwidths (14).

Decoding orientation from population activity. Since the variety of tuning modulations observed in single neurons pleads in favor of a multi-neuronal collaboration, we used a neuronal decoding method that probes for a population code 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, which is a probabilistic model that classifies data with multiple outcomes (28) (see Material and Methods). Here, this classifier was fed the fir-

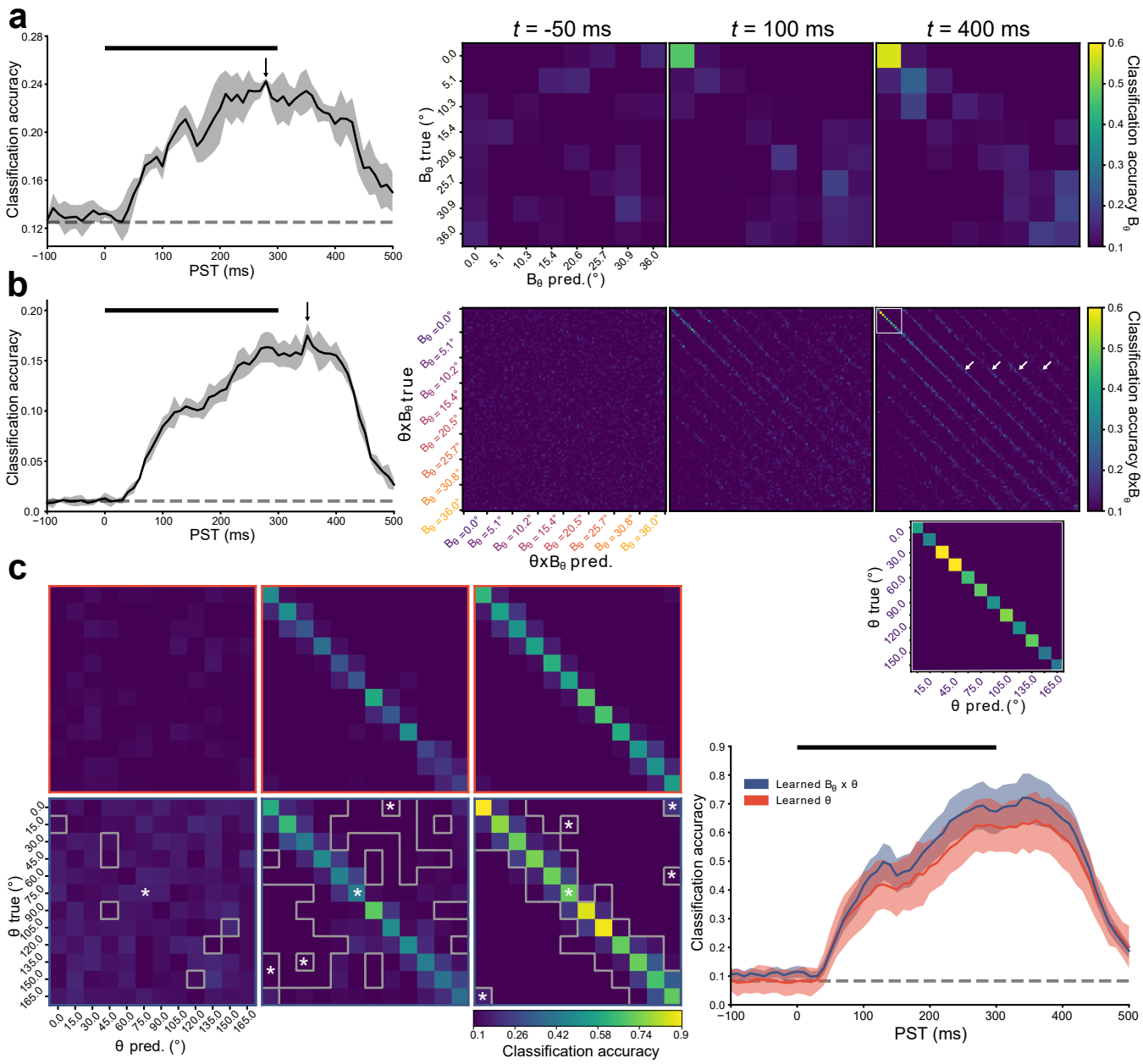


Figure 3. Decoding orientation bandwidth. (**a**) Left: Time course of a decoder trained to retrieve the orientation bandwidth (B_θ) of Motion Clouds. Solid dark line represent the mean accuracy of a 5-fold cross validation and filled contour the SD. The arrow points at the time at which maximum accuracy 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). 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 $\theta \times B_\theta$ decoder (as in b), 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).

ing rate of all recorded neurons in a given time window and learned a coefficient for each recorded neuron that predicts best the identity of a stimulus eliciting the population 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 with optimal parameters (Supplementary Figure 4a) . These were then able to predict the correct Motion Clouds' θ well above the chance level of 1 out of 12 values. 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 decoding performance is sustained up to 150 ms after the end of stimulation, indicating that late neural activity can serve to decode the orientation of stimuli for all B_θ (Figure 2b). 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 (number of significant $y_{pred} = y_{true}$ instances) of the $B_\theta = 0°$ decoder, observed as early as 100 ms after the stimulation onset,

is close to the diagonalization of the $B_\theta = 36.0°$ decoder observed hundreds of ms later (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, for all B_θ, albeit with different dynamics. The short delay required to perform decoding above chance level is congruent with the feedforward-like 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) reached 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 bi-modal like distributions of the tuning steepness in the population (Figure 1c). 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 $\theta \times B_\theta$ 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). In line with the previous θ decoders (Figure 2b), the confusion matrix of this $\theta \times B_\theta$ decoder showed greater accuracy for patterns of narrow bandwidth (Figure 2b inset, 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 (only elements within the main diagonals are correct decoding occurences, where $y_{pred} = y_{true}$). Since the addition of B_θ worsens the decoding of θ, even at $B_\theta = 0.0°$, 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 $\theta \times B_\theta$ 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 $\theta \times B_\theta$ decoder performed better than the B_θ-agnostic decoder (12/12 classes significant, $p < 0.05$, 1000 permutations

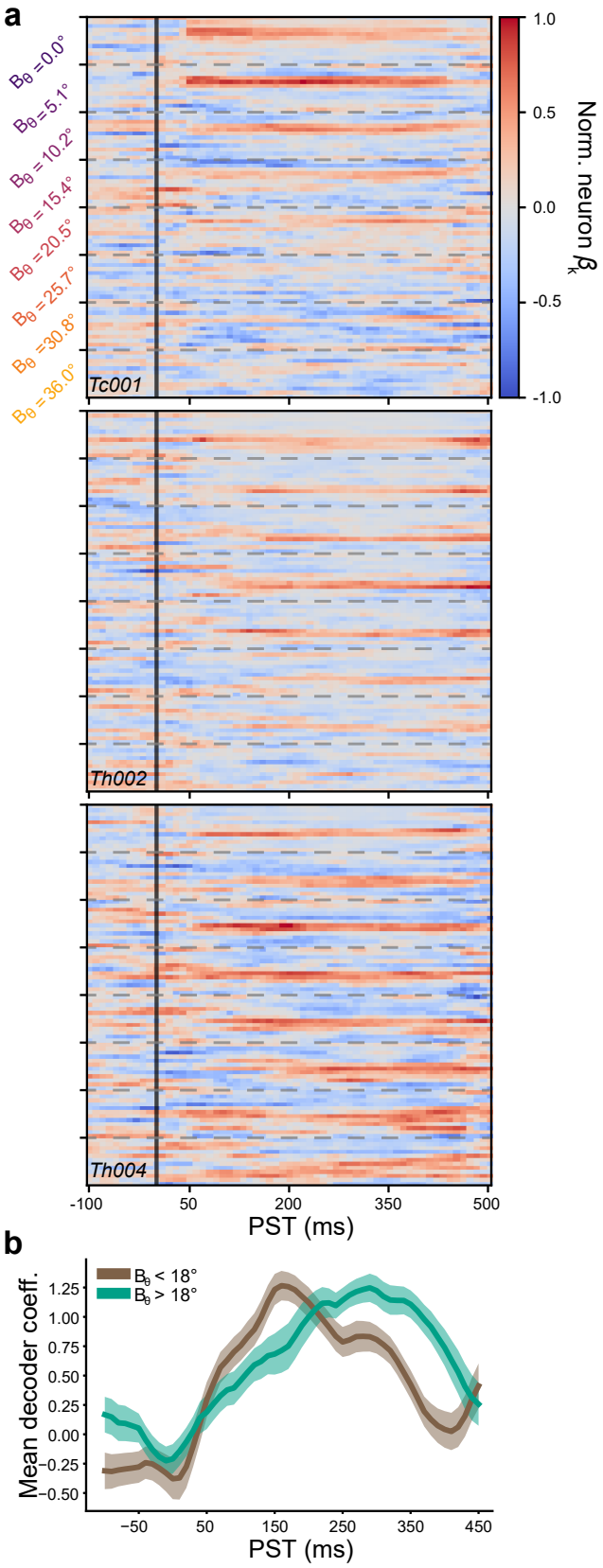


Figure 4. Neuronal contributions to population decoding. (**a**) Coefficient maps around the peri-stimulus time (PST) of three neurons. (**b**) Time course of the total coefficients of the $\theta \times B_\theta$ decoder, averaged on a 50 ms window. Darker lines indicate mean values and filled contour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

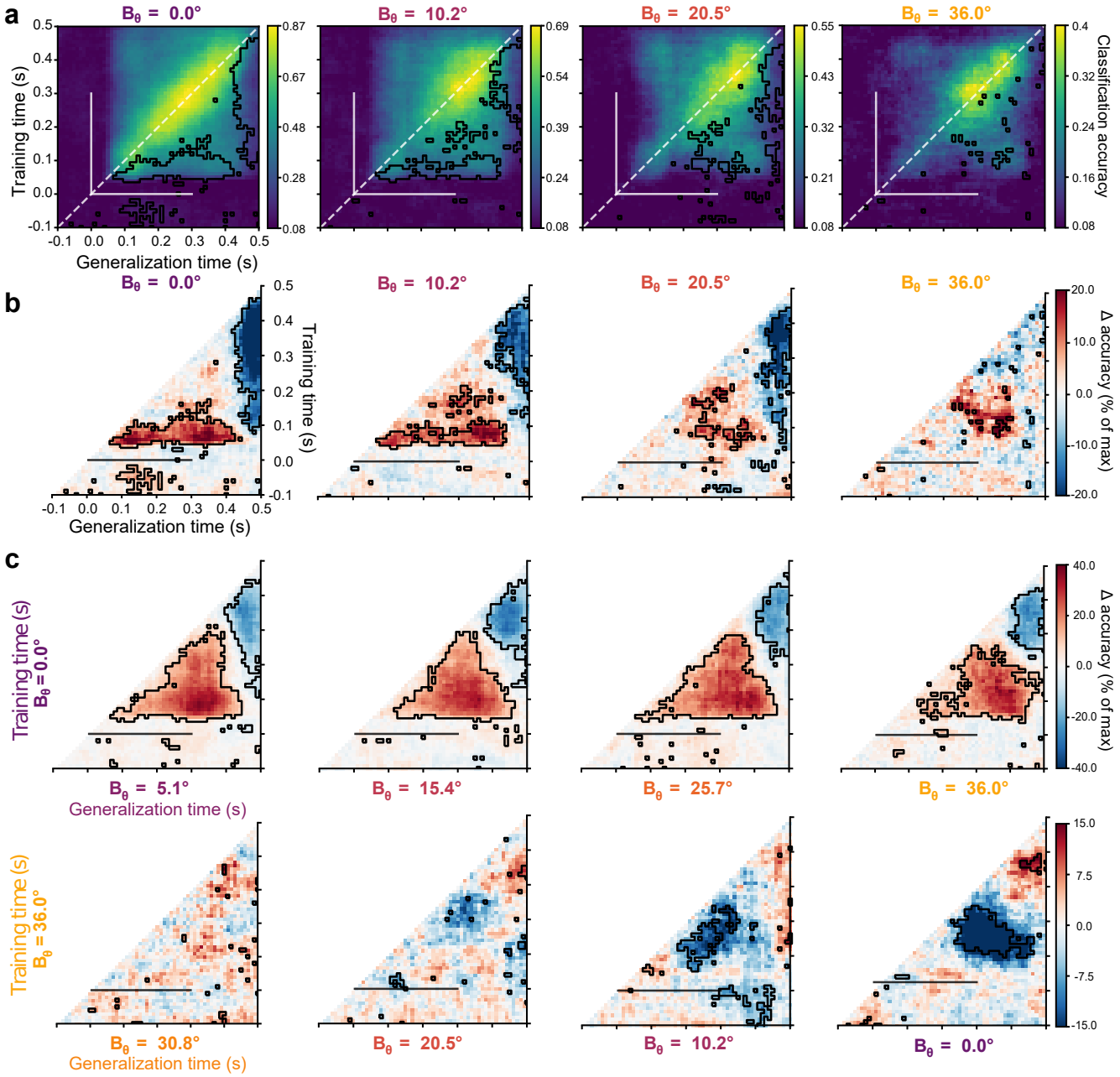


Figure 5. Temporal generalization. (**a**) Temporal generalization matrices of four independent orientation decoders trained for different bandwidths. Black contoured regions indicate significant difference (1000 permutations test, n=6, threshold $p < 0.05$) compared to the symmetrical region with respect 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 6b, in which a decoder is trained on the narrowest orientation bandwidth and generalized onto increasing B_θ (from left to right). Bottom: Generalizing from broadest to narrowest orientation bandwidths (from left to right).

test). Prior knowledge on the orientation bandwidth associated with an oriented input thus yielded better performances of $y_{pred} = y_{true}$ orientation decoding.

Temporal dynamics of bandwidth decoding. The visualization of the decoder's weight can give some indication as to how this dual $\theta \times B_\theta$ code is implemented in single neurons. The decoder learned for each neuron and at each time step a vector β_k of K elements. Representing all the independently-learned vectors through time creates a coefficient map that allowed to visualize the contribution by individual neurons to the population code in the $\theta \times B_\theta$ 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 different B_θ values. Similar to the behavior of tuning curves (Figure 1b), the coefficient maps showed for some neurons a decrease of the orientation selective decoding coefficients as B_θ increased (coefficient maps of neurons Tc001 and Th002), while others remained tuned yet with an increasing width and time delay (neuron Th004).

While the tuning curves were linked to the neuron's receptive

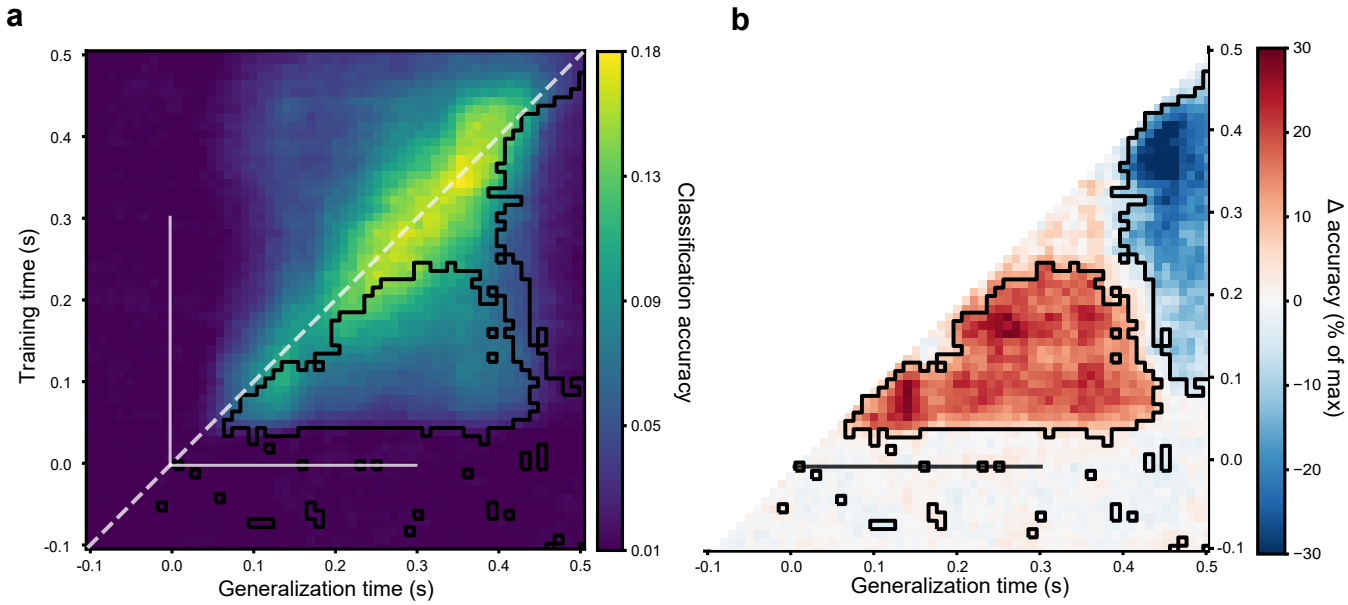


Figure 6. Temporal generalization of the orientation and orientation bandwidth ($B_\theta \times \theta$) decoder. White lines represent stimulation duration. Contoured regions indicate significant difference (1000 permutations test, n=6, threshold p < 0.05) compared to the symmetrical region with respect to the time identity line (dashed white). (**b**) Asymmetry of the temporal generalization matrix. Black line represent stimulation duration and black contours regions are of significant asymmetry.

field, the coefficient maps here are comparable to an "emitting field", i.e. to the content 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 $\theta \times B_\theta$ identity of a stimulus? A possible disambiguation of the stimulus' identity comes from the dynamics 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). Since the decoding of orientation and bandwidth evolved through trial time, we investigated the detailed dynamics of the neural code by using temporal generalization (29). Briefly, a decoder was trained and tested at two different time points, t_{train} and t_{test}, with the underlying hypothesis that a successful temporal generalization relies on the existence of a similar neural code at both training and testing times. More specifically, this amounts to use the coefficient β_{k1} learned at t_{train} in the example given in Figure 2a, and to decode the activity at time t_{test}. 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 $t_{\text{train}} \neq t_{\text{test}}$ allowed us to assess the unraveling of the neural code in time. Note that generalization points where $t_{\text{train}} = t_{\text{test}}$ (on the white dashed line, Figure 5a) are the same measurement as the time course of accuracy shown previously (Figure 2b). The accuracy decreased slowly when moving away from this time identity line, showing a good degree of decoding generalization throughout time. The upper halves of the temporal generalization matrices above the time identity line correspond to generalizations where $t_{\text{train}} \geq t_{\text{test}}$, i.e. generalizing backward in trial time, while generalizations below the time identity line, where $t_{\text{train}} \leq t_{\text{test}}$, are generalizations forward in trial time. Hence, subtracting the upper half from the lower half of each matrix measures the asymmetry around the time identity line (contoured clusters, Figure 5, Supplementary Figure 6) and gives an indication of the direction of the neural code generalization in the temporal dimension. A positive asymmetry (red clusters) indicates a better generalization forward in time, while a negative asymmetry (blue clusters) indicates a better backward generalization. For instance, the temporal generalization matrices of narrow B_θ-specific orientation decoders showed a significant asymmetry (Figure 5b), 44% significantly asymmetric transpositions, $B_\theta = 0.0°$, $p < 0.05$, 1000 permutations test corrected). 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 in Figure 5a, positive values in Figure 5b) is not identical to training and transposing it backwards in trial time (upper half Figure 5a, negative values in Figure 5b). Notably, this asymmetry is positive when $t_{\text{train}} \approx 100$ ms and $t_{\text{test}} \geq 100$ 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. In other words, narrow B_θ information is encoded in the early transient population activity and this neural code is kept the same for the entire trial. Conversely, the asymmetry became negative for $t_{\text{test}} \geq 350$ ms, independently of the trial time, implying that the late offset activity's structure is intrinsically different and cannot be decoded based on the earlier activity. 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

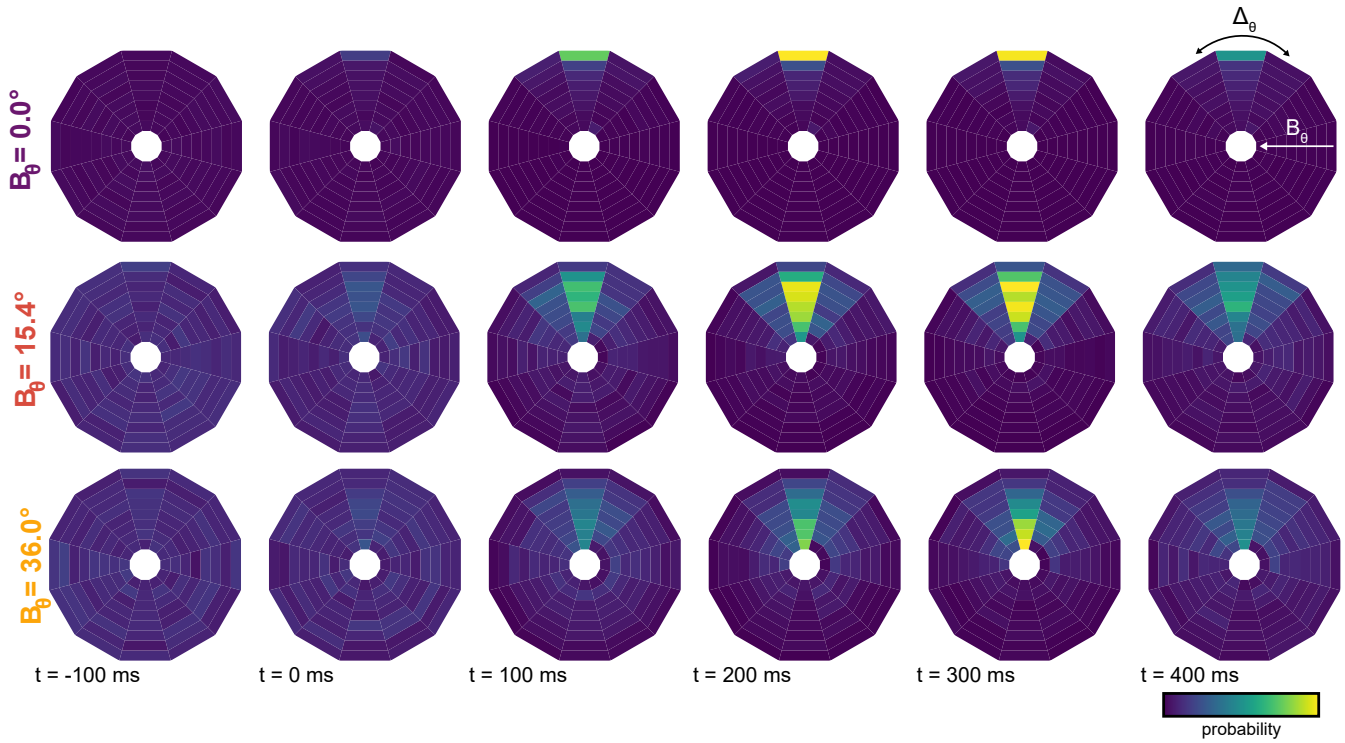


Figure 7. Dynamical convergence of the neural representations. The probability of the decoded stimuli is plotted as a function of the error on the θ identity of the stimulus, Δ_θ (represented as the angle of the polar representation with the $-90°$ to $90°$ range is unfolded over the circle and relative to the up vertical axis), while the eccentricity of each bin corresponds to its inferred B_θ (broader bandwidths toward the center). For stimuli of different B_θ, the temporal evolution of the probability from the decoder is normalized by row (with light values denoting high probabilities of decoding, and darker lower probabilities).

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 more suited to decode narrow B_θ, whereas the symmetry of the $B_\theta = 36.0°$ matrix implies that the late neural code for broad stimuli is always present but rises 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 6b). 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 t_{train} while negative asymmetry is still present in late t_{train} shows that a narrow B_θ code is always present at the transient of the stimulation and stays similar across time points and B_θ. 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 $B_\theta = 36.0°$ to $B_\theta = 0.0°$, $p < 0.05$, 1000 permutations test). Namely, a pattern of negative asymmetry was present for $t_{\text{train}} \approx 150$ ms and $t_{\text{test}} \geq 250$ ms, which corresponds to the region of non-significant asymmetry in 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 dynamics. The transient activity evoked by a narrow stimulus bandwidth generates a code that can be generalized to later time (and larger bandwidths), while the broad bandwidth codes can be generalized independently of time and bandwidth, but only dominates the global population code at later times.

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 $\theta \times B_\theta$ 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 broad B_θ code. Hence, the overall $\theta \times B_\theta$ population code retains the specificity of the bandwidth-dependent dynamics.

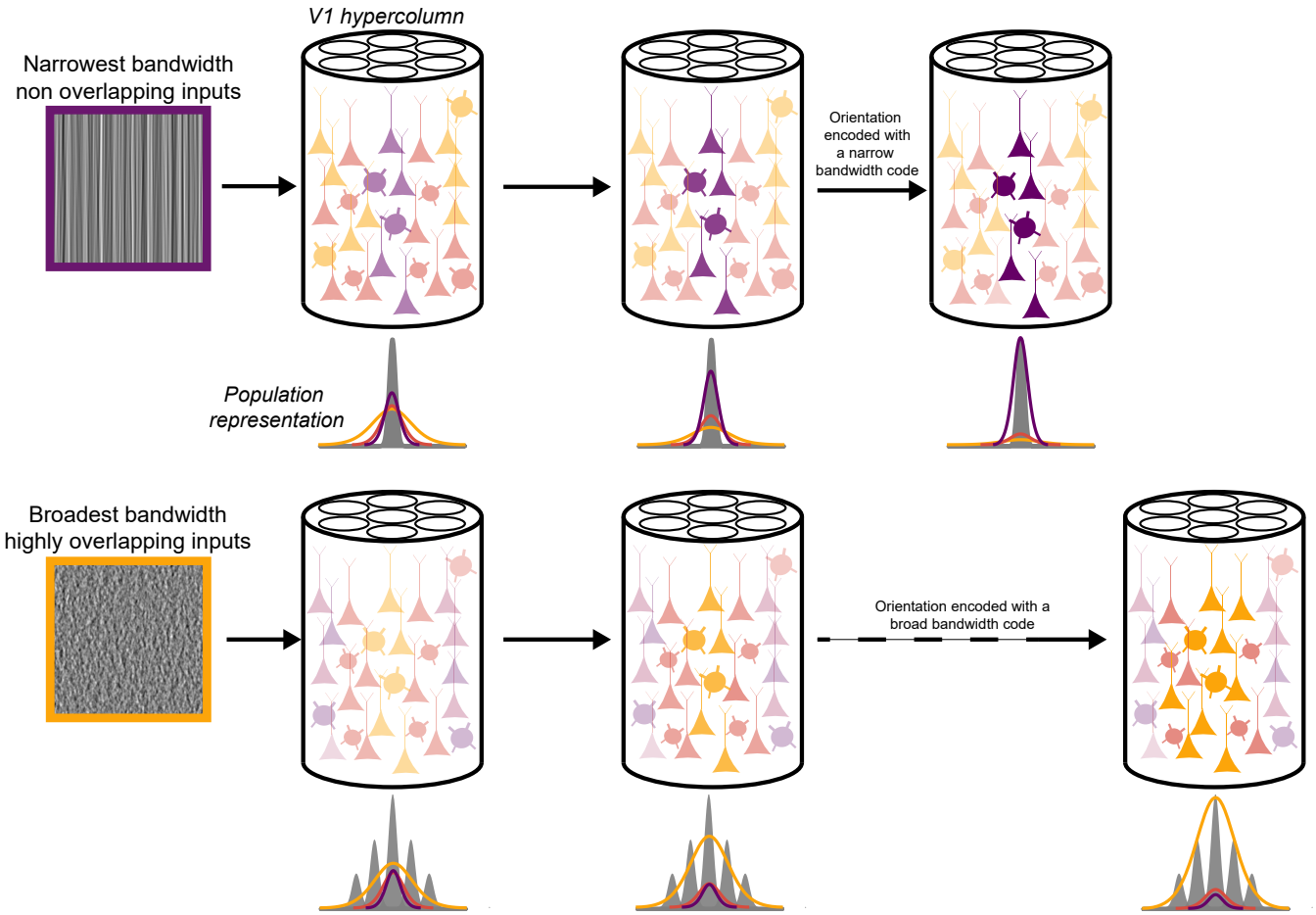


Figure 8. Schematic processing of orientation bandwidth within a cortical hypercolumn, which contains all possible oriented columns. Neurons' colors represent the nature of their contribution to the bandwidth code, while the intensity of the color represents the degree of this contribution. The overall representation of the population is represented in relation to the thalamic input, in grey. In the case of a narrow bandwidth Motion Clouds, the orientation content is highly concentrated and can be easily read out by V1 with classical orientation processing means. For largest bandwidths, the distribution of input orientations is enlarged and retrieving the orientation of the stimulus requires separating the distributions into multiple components that can be each orientation-processed, implicating additional cortical processing time while reducing the maximum accuracy of orientation decoding. As the processing time increases, a consensus on the bandwidth emerges in the population.

What is the overall dynamic of the population code ? The probabilistic output of the decoder (Figure 7) hints at a separate processing for each of the different B_θ codes. Indeed, for narrow bandwidth, the most probable pattern is correctly identified right at the onset of the stimulus presentation and stays stable through time. Furthest errors of orientation decoding, Δ_θ are on the order of one step in our stimulation parameters (15°). For broader B_θ values, there is a no clear consensus on the θ of the stimulation at the onset, but the correct identity becomes apparent much later, as a gradient centered on the correct orientation and bandwidth emerges. This progressive convergence onto the correct stimulus' identity can be interpreted in terms of a segregation of oriented inputs within cortical populations (Figure 8). In that sense, the observed bandwidth code is a neural trace of the undergoing separation of orientations. Neurons would contribute to a population "vote" on orientation and bandwidth, which explains the stability of early, narrow bandwidth activity as a global consensus towards the most likely $\theta \times B_\theta$. The additional processing time required for stimulations of broader bandwidths is explained by the potential decomposition of broad Motion Clouds into segregated distributions whose orientations can be retrieved.

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 the orientation bandwidth, as well as a correlation between the steepness of these B_θ modulations and the tuning of the neuron, evocative of a relationship between tuning width and stimulus (30). This heterogeneously tuned population, both in θ and B_θ, is suited to decode orientation mix-

tures of varying bandwidths and is known to be pivotal for encoding natural images in V1 (31). We performed a biologically plausible readout (24, 32) on this population activity and found that for narrow B_θ stimuli, above chance level decoding was performed in a timing coherent with feedforward-like processing. More intriguing, the delay required to infer the identity of stimuli with maximum accuracy, especially for broader B_θ, would suggest the involvement of a slower neuronal pathway.

Finding a suitable origin for this late response is easier when examining the study of orientation bandwidth within a predictive coding framework. In predictive coding, predicted and actual sensory state generate prediction errors to update the internal state of the generative model of the world. These predictions are modulated proportionally to their precision, such that highly precise predictions errors on sensory input cause a greater update of the internal model. Now, considering that the bandwidth of our stimuli are linked to the inverse variance of their orientation distributions, that is, their precisions, Motion Clouds offer an ideal tool to investigate the role of precision-weighting of the sensory inputs. In predictive coding, the top-down inference on sensory states are classically assigned to feedback connectivity (33), which is the first putative candidate for the neural substrate of precision modulation. This would imply that the low and high precision codes co-exists temporally but not spatially, and correspond 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 (34) but provides contextual modulations of V1 (35), a suitable synonym for precision weighting (36).

Local cortical connectivity in V1 is also likely involved. Recurrent connectivity implements complex computations (37) and maintains tuning to features that cannot be encoded by single neurons (38), making it relevant in processing stimuli such as Motion Clouds. 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 (39). If the implementation of precision weighting comes from this neural pathway, then the increased delay to reach maximum accuracy for broad bandwidth stimuli would stem from multiple iterations of computations implemented by recurrent interactions, which is coherent with the speed of slowly-conducted horizontal waves in the orientation map (40). 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 (41). This inhibitory process would be well suited to the interpretation of our results in terms of separation of the multiple orientations (Figure 8), which would be performed by local cortical competition.

Whether feedback or recurrent, both these interpretations of the results fit the view that the precision of feedforward predictions errors can be assigned to supragranular neurons (42). 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 (36).

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 (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, which could emerge independently throughout the visual hierarchy. Within V1, the non-simultaneity of our electrophysiological recordings prevents us from further interpretations on the mechanisms involved in precision processing. Previous publications have reported population decoding by merging neural activity across electrodes or experiments (46, 47), which we validated in our data (Supplementary Figure 5).

Naturalistic stimuli have been used here to shed light on the processing of orientation precision. This problem was framed in the form of an instantaneous snapshot of visual inputs and can now be transposed to natural images and their exploration with saccadic behaviors, which adds a temporal dimension to the orientation distributions (48). Regardless of the involvement of saccades, the functional organization of precision processing in V1 remains an open question. The substrate for cortical dynamics on the scale of hundreds of milliseconds is still ill-defined but, as previously mentioned, slow temporal events are likely related large-scale neural activity. Naturally, one can then wonder how precision is processed at the mesoscale level. Could the nonlinearities observed be organized in maps in the primary visual cortex and if so, how would they relate to the maps of orientation?

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 CO_2 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 (49). 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 (50). Single units were isolated using Kilosort spike sorting software (51) and manually curated using the phy software (52). 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 (53) 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 are Motion Clouds (26), a class of band-pass filtered white-noise textures (54) 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 $f_r = \sqrt{f_x^2 + f_y^2}$ is the radial frequency, 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 (55). 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 (56) 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) (57) 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 (58).

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 $\arg\max_{B_\theta=0.0^\circ}$ 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 $B_\theta = 0.0^\circ$ and $B_\theta = B_{\theta_{\text{untuned}}}$.

Population Decoding. The stimulus identity was decoded using a multinomial logistic regression classifier (28), which can be interpreted as an instantaneous, neuronal plausible readout of the population activity (24). For a given stimulus, 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 (28) 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:

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

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 4).

Decoding accuracy was reported as the average accuracy across classes, also known as the balanced accuracy score (59), 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 (29). 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) and corrected for multiple comparisons.

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 and Jonathan Vacher for fruitful exchanges on the formalization of the generation of synthetic images and for his contributions to the analysis of neurophysiological recordings. 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 (60), scikit-learn (61), numpy (62), PyTorch (63), lmfit (64) and Matplotlib (65).

AUTHOR CONTRIBUTIONS

L.U.P., C.C. F.C., N.C., and H.J.L. designed the study. H.J.L., N.C. and L.K. collected the data. H.J.L., N.C. 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

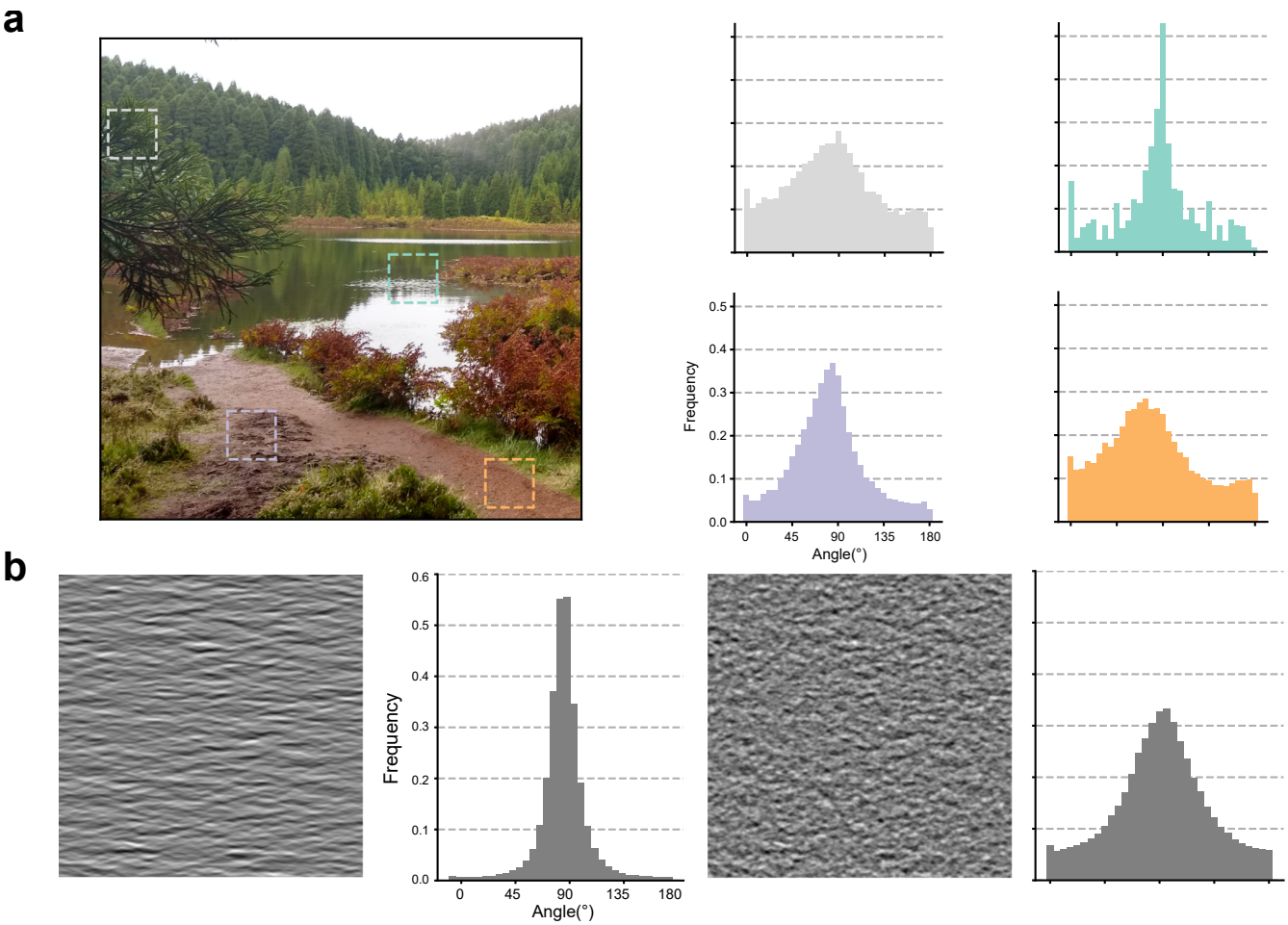
The authors declare no competing financial interests.

References

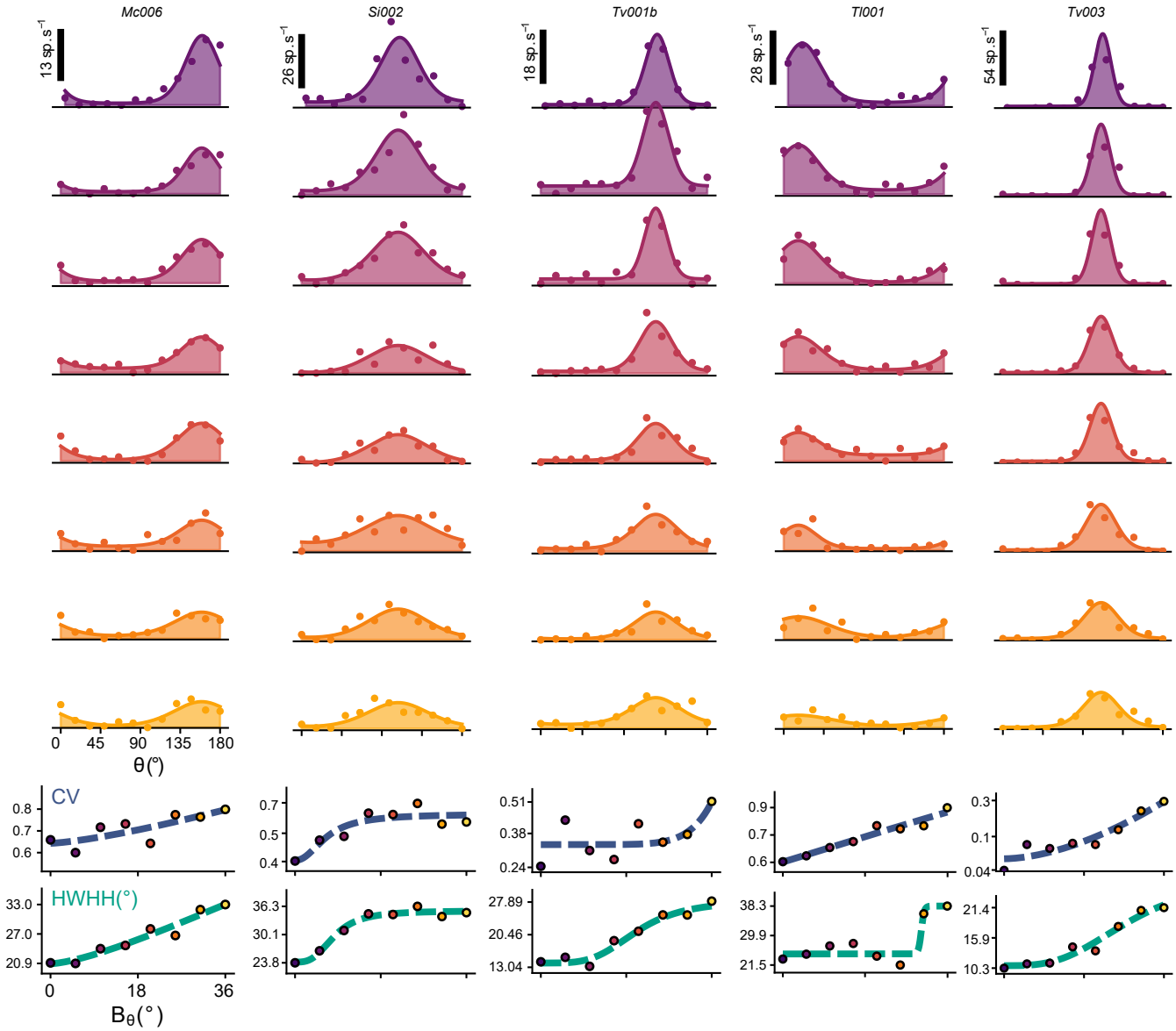
1. Hubel, D. H. & Wiesel, T. N. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology* **148**, 574 (1959).
2. Priebe, N. J. Mechanisms of orientation selectivity in the primary visual cortex. *Annual review of vision science* **2**, 85–107 (2016).
3. Fiser, J., 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. Julesz, B. Textons, the elements of texture perception, and their interactions. *Nature* **290**, 91–97 (1981).
18. Phillips, G. C. & Wilson, H. R. Orientation bandwidths of spatial mechanisms measured by masking. *JOSA A* **1**, 226–232 (1984).
19. Heeley, D., Timney, B., Paterson, I. & Thompson, R. Width discrimination for band-pass stimuli. *Vision research* **29**, 901–905 (1989).
20. Heeley, D. W. & Buchanan-Smith, H. M. The influence of stimulus shape on orientation acuity. *Experimental Brain Research* **120**, 217–222 (1998).
21. Keeble, D., Kingdom, F., Moulden, B. & Morgan, M. Detection of orientationally multimodal textures. *Vision Research* **35**, 1991–2005 (1995).
22. Beaudot, W. H. & Mullen, K. T. Orientation discrimination in human vision: Psychophysics and modeling. *Vision research* **46**, 26–46 (2006).
23. 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).
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. Bishop, C. M. *Pattern recognition and machine learning* (springer, 2006).
29. King, J.-R. & Dehaene, S. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in cognitive sciences* **18**, 203–210 (2014).
30. Zhang, K. & Sejnowski, T. J. Neuronal tuning: To sharpen or broaden? *Neural computation* **11**, 75–84 (1999).
31. 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).
32. Schwartz, O., Pillow, J. W., Rust, N. C. & Simoncelli, E. P. Spike-triggered neural characterization. *Journal of vision* **6**, 13–13 (2006).
33. 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).
34. 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).
35. Ito, M. & Gilbert, C. D. Attention modulates contextual influences in the primary visual cortex of alert monkeys. *Neuron* **22**, 593–604 (1999).
36. 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).
37. Chance, F. S., Nelson, S. B. & Abbott, L. F. Complex cells as cortically amplified simple cells. *Nature neuroscience* **2**, 277–282 (1999).
38. Benvenuti, G. *et al.* Scale-invariant visual capabilities explained by topographic representations of luminance and texture in primate V1. *Neuron* **100**, 1504–1512 (2018).
39. 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).
40. 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).
41. Carandini, M. & Ringach, D. L. Predictions of a recurrent model of orientation selectivity. *Vision research* **37**, 3061–3071 (1997).
42. Bastos, A. M. *et al.* Canonical microcircuits for predictive coding. *Neuron* **76**, 695–711 (2012).
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. 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).
50. Siegle, J. H. *et al.* Open ephys: an open-source, plugin-based platform for multichannel electrophysiology. *Journal of neural engineering* **14**, 045003 (2017).
51. 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).
52. Rossant, C. *et al.* Spike sorting for large, dense electrode arrays. *Nature neuroscience* **19**, 634–641 (2016).
53. Peirce, J. *et al.* Psychopy2: Experiments in behavior made easy. *Behavior research methods* **51**, 195–203 (2019).
54. Carandini, M. & Heeger, D. J. Normalization as a canonical neural computation. *Nature Reviews Neuroscience* **13**, 51–62 (2012).
55. 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).
56. 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).
57. Swindale, N. V. Orientation tuning curves: empirical description and estimation of parameters. *Biological cybernetics* **78**, 45–56 (1998).
58. 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).
59. 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).
60. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272 (2020).
61. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
62. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272 (2020).
63. 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).
64. Newville, M. *et al.* Lmfit: Non-linear least-square minimization and curve-fitting for python. *ascl* ascl–1606 (2016).
65. Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering* **9**, 90–95 (2007).

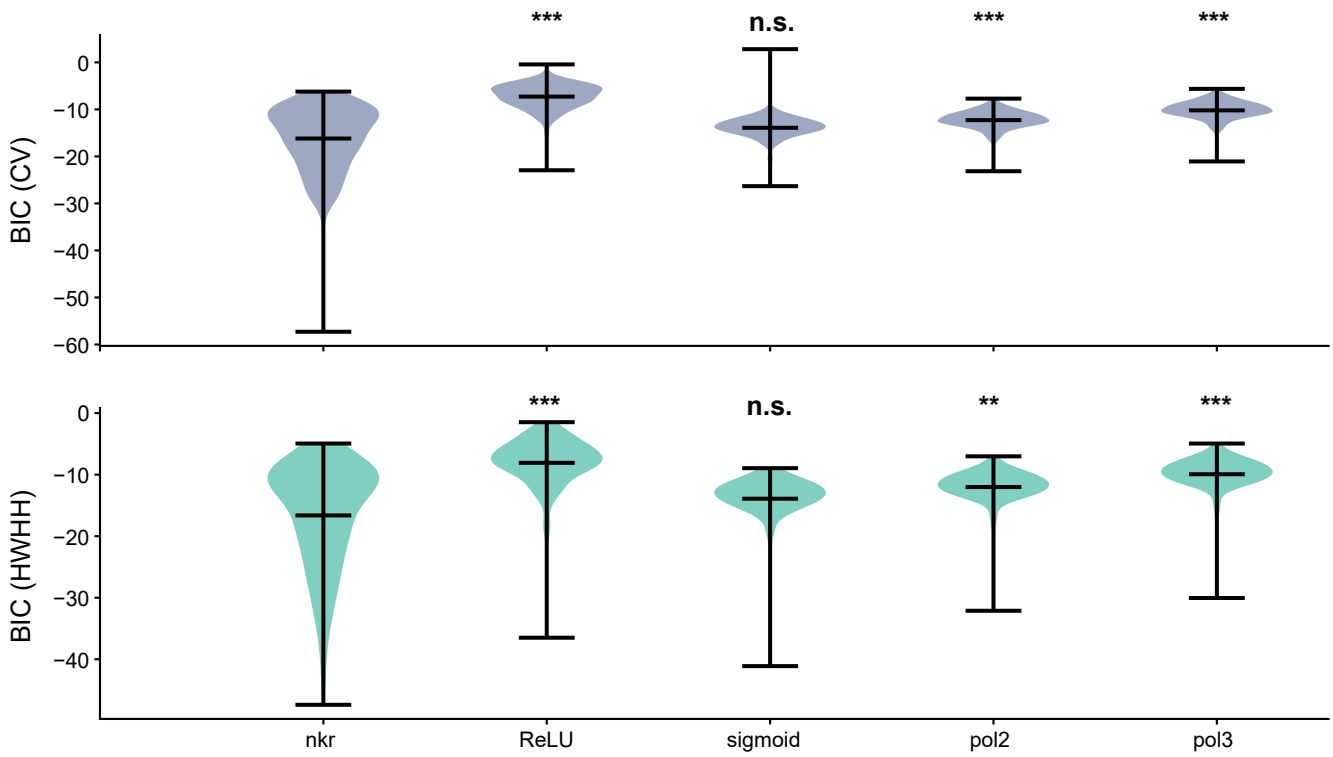
Supplementary Figures



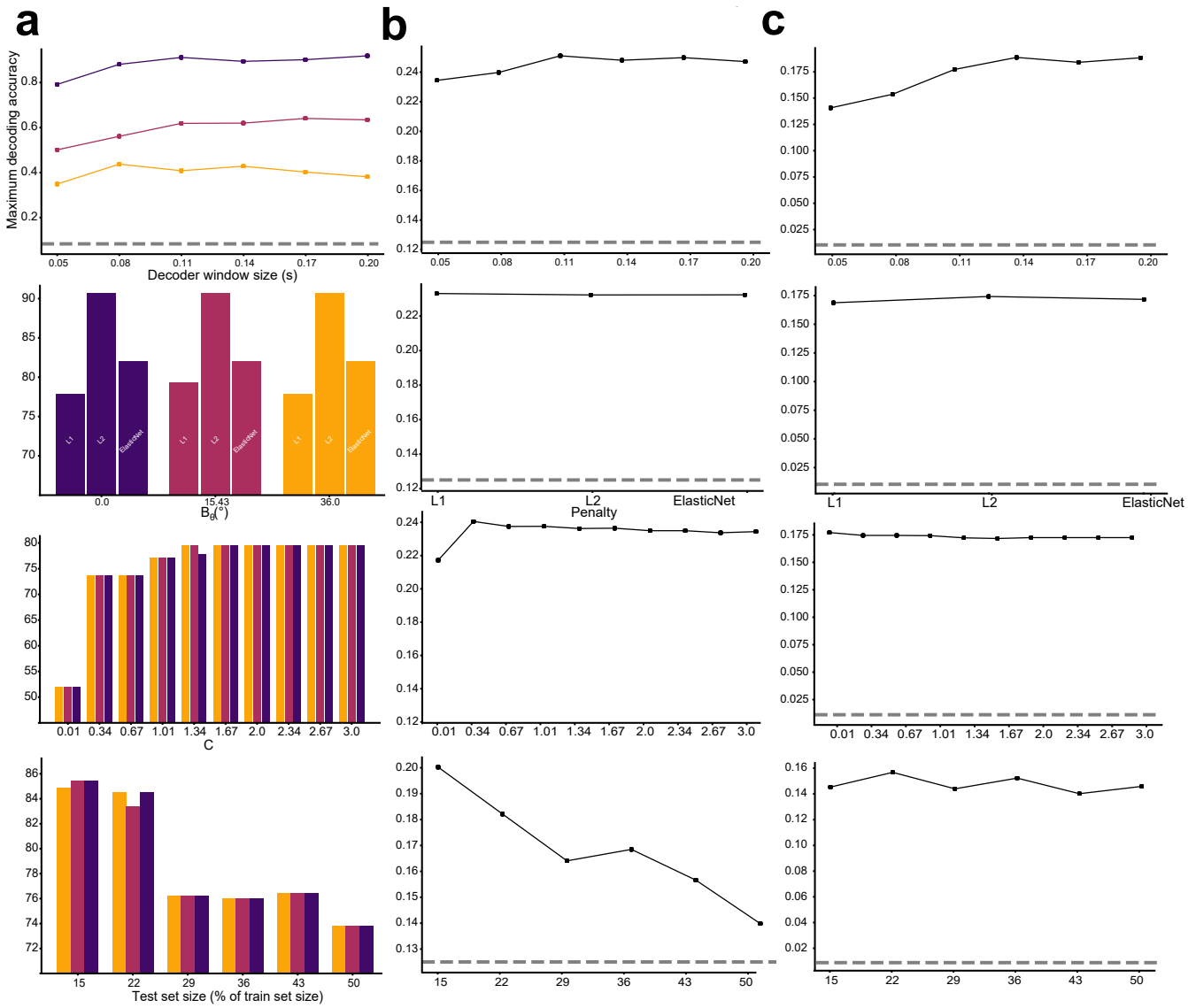
Supplementary Figure 1. Motion Clouds capture the variety of orientation distributions found in natural images. (**a**) Distribution of orientations found in a natural image (São Miguel Island, Azores, Portugal, picture taken by H.J.L.), retrieved with a Histogram of Oriented Gradients (16x16 cell size, 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 $B_\theta = 36.0°$.



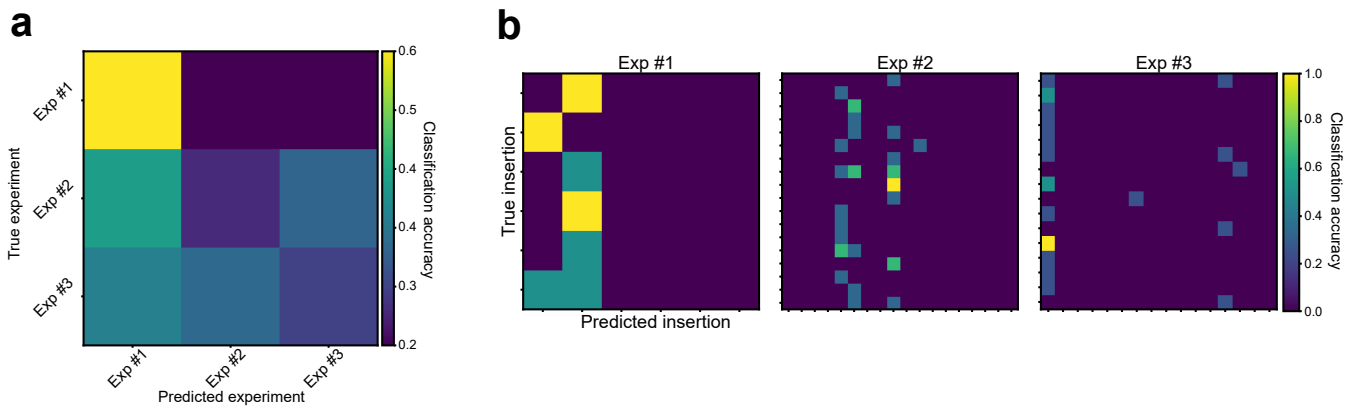
Supplementary Figure 2. Additional examples of single neuron responses. Top: Tuning curves of 5 additional neurons in response to Motion Clouds of decreasing B_θ (top to bottom). Bottom: Changes in Half-Width at Half-Height (HWHH) and Circular Variance (CV) as a function of B_θ for the respective neurons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



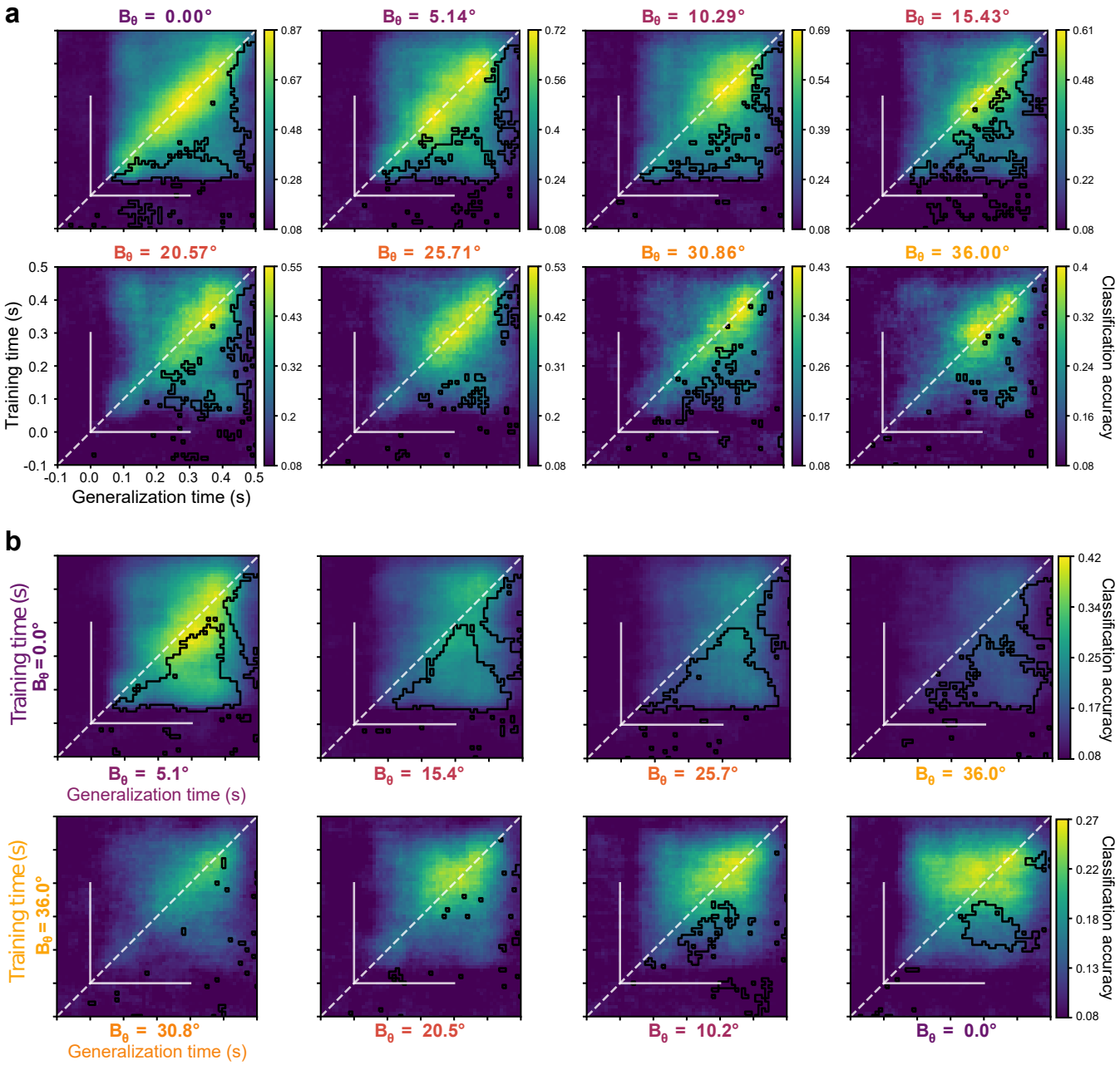
Supplementary Figure 3. Criterion on the model selection for HWHH and CV curves. Top: Violin plot of the Bayesian Information Criterion (BIC) of the CV curves of all recorded neurons. Each violin plot represent a different kind of fitted equation, respectively: Naka-Rushton (`nkr`, see Materials and Methods) ; Rectified Linear Unit (`ReLU`, $f(x) = max(0,x)$) ; logistic function (`sigmoid`, $f(x) = \frac{e^x}{e^x+1}$) ; second degree polynomial function (`pol2`, $f(x) = ax^2 + bx + c$) and third degree polynomial function (`pol3`, $f(x) = ax^3 + bx^2 + cx + d$). 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; $\star$, p < 0.05; $\star\star$, p < 0.01; $\star\star\star$, p < 0.001 (Kruskal-Wallis H-test, post-hoc Dunn Pairwise test, Bonferroni corrected).



Supplementary Figure 4. 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 5. 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 6. 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.