Choice can be predicted from populations of bursting neurons in superficial layers of monkey V1

Veronika Koren [1,2,4,*], Ariana R. Andrei [3], Ming Hu[3,5], Valentin Dragoi [3,*], Klaus Obermayer [1,2,6,*]

1 Institute of Software Engineering and Theoretical Computer Science, Technische Universität Berlin, 10587 Berlin, Germany **2** Bernstein Center for Computational Neuroscience Berlin, Germany **3** Department of Neurobiology and Anatomy, University of Texas Medical School, Houston, Texas, 77030, US **4** Present address: Institute of Mathematics, Technische Universität Berlin, 10623 Berlin, Germany **5** Present address: Baylor College of Medicine, Houston, TX 77030, US **6** Lead author

[*] Correspondence: valentin.dragoi@uth.tmc.edu, veronika.koren@tu-berlin.de, klaus.obermayer@tu-berlin.de

Highlights

- The choice can be predicted from the spiking activity in the cortical column of V1 of the macaque.
- The information on choice and on stimuli is partially overlapping.
- Bursty neurons in the superficial layer of the cortex are the principal carriers of the choice signal.
- Correlated spike timing between neurons with similar decoding selectivity helps encoding.

Abstract

Primary visual cortex (V1) is absolutely necessary for normal visual processing, but whether V1 encodes upcoming behavioral decisions based on visual information is an unresolved issue, with conflicting evidence. Further, no study so far has been able to predict choice from time-resolved spiking activity in V1. Here, we hypothesized that the choice cannot be decoded with classical decoding schemes due to the noise in incorrect trials, but it might be possible to decode with generalized learning. We trained the decoder in the presence of the information on both the stimulus class and the correct behavioral choice. The learned structure of population responses was then utilized to decode trials that differ in the choice alone. We show that with such generalized learning scheme, the choice can be successfully predicted from spiking activity of neural ensembles in V1 in single trials, relying on the partial overlap in the representation between the stimuli and the choice. In addition, we show that the representation of the choice is primarily carried by bursting neurons in the superficial layer of the cortex. We demonstrated how bursting of single neurons and noise correlations between neurons with similar decoding selectivity helps the accumulation of the choice signal.

Keywords: Visual cortex; cortical column; decoding; stimulus; choice; neural network; population code; information; discrimination; generalized learning

Introduction

In real-world scenarios, animals frequently make decisions based on the observation of their environment. In the context of such decision-making, it has been understood that encoding of sensory information and utilizing this information for guiding the behavior do not necessarily overlap (Panzeri et al. 2017). So far, the primary visual cortex has been shown to be a site of relatively sophisticated sensory processing, from which other brain areas read-out the sensory information in order to perform further computations, such as decision-making based on sensory cues. Corroborating this view has been the evidence of the choice probability in the middle and high visual cortical areas, as "early" as V2, but not in V1 (Nienborg and Cumming 2006). Recent study, however, has found the choice probability in V1 during a coarse orientation-discrimination task with artificial stimuli (Nienborg and Cumming 2014), arguing that, in order to observe choice probability, neurons must be topologically organized with respect to the task feature.

Representation of the behavioral choice in absence of sensory evidence has been extensively studied for three decades (Shadlen and Newsome 1996; Britten et al. 1996; Shadlen et al. 1996; Logothetis and Schall 1989; Roitman and Shadlen 2002; Nienborg and Cumming 2009). Studies on choice probability have primarily addressed the time-averaged activity of single neurons and their relation to the behavioral choice of the animal, and single neurons have been shown to correlate with the choice behavior in a variety of brain areas (Shadlen and Newsome 1996; Britten et al. 1996; Logothetis and Schall 1989; Uka and DeAngelis 2004; Roitman and Shadlen 2002). While it is interesting that there are single neurons in sensory areas whose activity predicts the choice with better-than-chance probability, it is unlikely that in the brain, the activity of a single neuron can be isolated from a vast number of cells that activate in parallel during a decision-making task (Nienborg et al. 2012). Recent years have seen the revival and the development of the idea that behaviorally relevant variables are encoded by collective dynamics of neural ensembles, evolving in real time (see, e.g., (Gao and Ganguli 2015; Saxena and Cunningham 2019; Koren and Denève 2017)). While a major obstacle for testing theoretical predictions on experimental data is the incompleteness of datasets, recent analysis of a nearly complete networks in the retina demonstrates that a good description of neural activity has to take into account single neuron firing rates, the strength of pairwise interactions and a global modulation of the network activity ((Tkačik et al. 2014)). Population codes seem to be appropriate to study neural responses in the auditory (Harris et al. 2011; Luczak et al. 2013), motor (Churchland et al. 2012), prefrontal (Rigotti et al. 2013), and visual cortex (Ni et al. 2018).

In the present study, we ask if the behavioral choice can be predicted from the spiking activity of neural populations in V1. We design a time-resolved code that reflects the computation of the low-dimensional choice variable in animal's brain. Neural computation of the choice presumably occurs through spikes of neural populations and takes place in real time to contribute to animal's behavior as it unfolds. It is therefore important to study time-resolved representations of the choice. Moreover, we use naturalistic stimuli that are close to psychophysical threshold. The requirement of the time-resolved population code, and the use of naturalistic stimuli close to the psychophysical threshold have, to our not knowledge, not yet been shown to yield a prediction of the choice in V1. We show that the choice cannot be predicted with classical decoding procedures, but requires the extension of classical decoding schemes.

In the experimental setup of the current study, animals sequentially view pairs of naturalistic visual stimuli, the target and the test stimulus (Figure 1A). The two stimuli can either be the same (condition Match) or not (condition Non-match), where the only change in non-matching condition is the change in the orientation of the test stimulus (Shahidi et al. 2019). After the visualization of the two stimuli, the animal reports the choice ("same" or "different") by holding or releasing a bar. We can consider two classification problems, one where classification relies on the information on both the stimulus and the choice (variable $stimulus + choice$), and another, where the classification relies on the information about the choice alone (variable *choice*, Figure 1B). Note that classification of $stimulus + choice$ relies on correct behavior while one of the conditions in *choice* is associated with incorrect behavior. In our previous work, we utilized classical decoding procedure and found that spike counts in V1 successfully predict the binary variable $stimulus + choice$ (Koren et al. 2020).

In another previous work, we decoded *stimulus* + *choice* as a low-dimensional representation of spike trains without temporal averaging, and found that the representation is relying mostly on the activity of neurons in the superficial cortical layers (Koren et al. 2019). Stimulus categories Match and Non-match (see Figure 1B) might be represented as contextual cues (Muckli et al. 2015; Zipser et al. 1996; Juan and Walsh 2003), possibly through the integration of bottom-up and top-down signals Larkum (2013).

In this study, we address decoding of the variable *choice* with a time-resolved population code. We find that the variable *choice* can be predicted through generalization of learning of the neural representation (Bernardi et al. 2020). In particular, we utilize a generalization of learning from *stimulus*+*choice* to *choice*. We implement such generalization by training an optimal linear decoder in the context of *stimulus* + *choice*, and applying the structural features to spike trains in single trials, thus constructing a real-time low-dimensional representation of the variable *choice*. Note that we use the term "representation" in a broad sense - if a variable can be decoded from neural activity, we say that neural activity represents the variable. If the representation in the context of *stimulus* + *choice* relies exclusively on the difference in stimuli, learning to discriminate the variable *stimulus* + *choice* cannot be generalized to the discrimination of the variable *choice*. If, on the contrary, the representation of the choice and the stimulus classes in V1 are mixed (Rigotti et al. 2013), learning in the context of *stimulus* + *choice* can possibly generalize to the representation of *choice*, due to the partial overlap in the information between the two classification problems (Panzeri et al. 2017). Note that mixed representations have been recently demonstrated in a variety of cortical areas, including V1 (Stringer et al. 2019).

In the present work, we show that, contrary to previous beliefs, upcoming behavioral choice can be decoded from the spiking activity in V1 in a time-resolved population code. We show that the choice related information is mostly present in superficial layers of the cortex and primarily carried by bursty neurons. Finally, we suggest how bursting and correlated spiking of neurons with similar decoding selectivity helps the accumulation of the time-resolved choice signal.

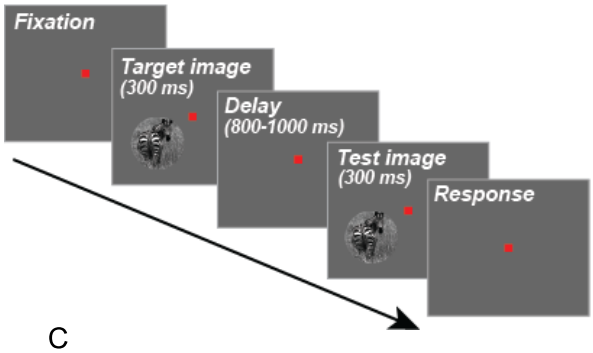
Results

Behavioral choice can be decoded from V1 spiking activity through generalized learning of neural representations

Two adult male macaques were trained on a delayed match-to-sample visual task. The subject visualized the target and the test stimuli, with a delay period in between (Figure 1A). The target and the test stimulus were either identical (condition match) or not (condition non-match), with the only possible difference being the change in the orientation of the test stimulus. In condition non-match, the angle of rotation of the test stimulus is random and can be clockwise or anti-clockwise. The task of the subject was to communicate its decision about the stimulus class ("same" or "different", see methods). The stimuli were complex naturalistic images in black and white, depicting an outdoor scene, and their identity changed from one recording session to another. The multiunit signal has been captured with linear arrays, inserted perpendicularly to the cortical surface and recording the activity across the cortical depth. After spike sorting, we obtained on the order of 10 units in each recording session, and the total of 160 units in 20 recording sessions. Example spike rasters for three simultaneously recorded neurons, belonging to the superficial (supragranular or SG), middle (granular or G) and deep (infragranular or IG) cortical layer, are shown on Figure 1C.

Single neuron statistics (Figure 1D,E) and standard multivariate decoding methods from populations of simultaneously recorded neurons (Figure 1F) were not predictive of the choice. Firing rates of single neurons varied strongly across units, but were similar for the choice "same" and "different" (Figure 1D). Coefficient of variation of single neurons (CV_2, see methods) varied around 1, which is the CV_2 of the Poisson process, and was strongly positively correlated across conditions (Figure 1E). Prediction of the *choice* with a linear decoder utilizing parallel spike counts was not significantly different from chance (average prediction accuracy of 0.51; permutation test with $N_{perm} = 1000$ random permutations without repetition; Figure 1F, black crosses). Note that random permutations are technical replicates (Blainey et al. 2014).

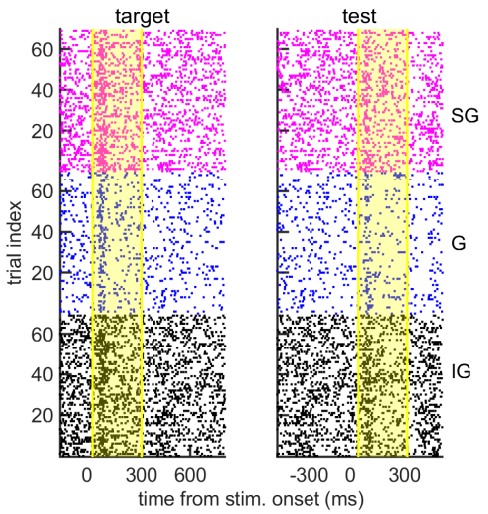
A



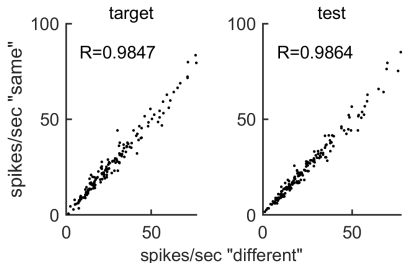
B

Condition	Stimulus	Choice	Info. content
correct non-match	non-match	"different"	stimulus +
correct match	match	"same"	choice
correct non-match	non-match	"different"	choice
incorrect non-match	non-match	"same"	

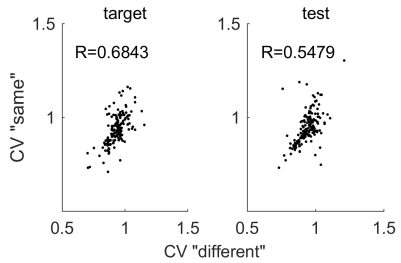
C



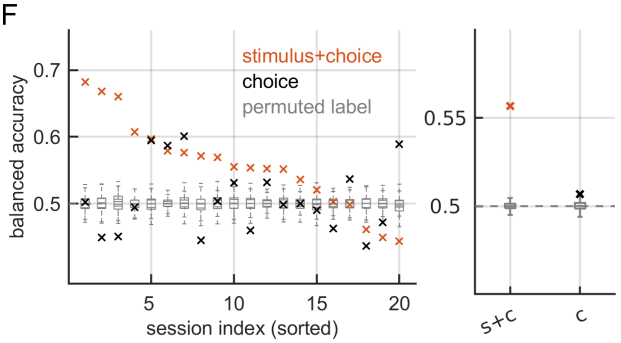
D



E



F



G

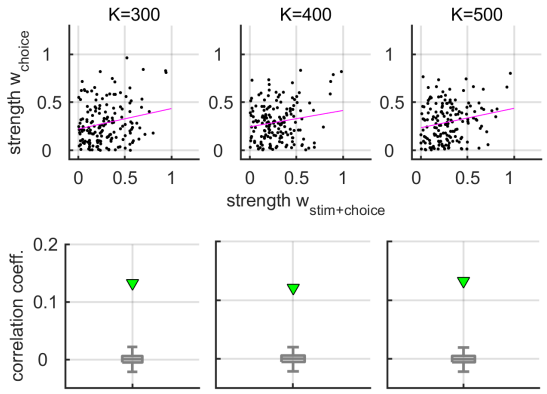


Figure 1. The structure of population responses in *stimulus + choice* **and** *choice* **is partially overlapping. (A)** Informational content of two classification problems. **(B)** Experimental paradigm. The subject visualizes the target and the test stimulus, with a delay period in between. **(C)** Spike rasters of three representative single neurons for the behavioral choice "different". We show spike rasters of three neurons, one in superficial layer of the cortex (SG, magenta), one in the middle layer (G, blue) and one in the deep layer (IG, black). **(D)** Scatter plot of the mean firing rate for the choice "different" (x-axis) and for the choice "same" (y-axis). We used the time window [0,400] ms w.r.t. the stimulus onset. R is the Pearson correlation coefficient. **(E)** Same as in **(D)**, but showing the coefficient of variation with time-dependent firing rates (CV_2). **(F)** Prediction accuracy of a linear classifier for the classification problem of *stimulus + choice* (red) and *choice* (black). Left plot shows results in recording sessions and right plot shows session-averaged results. Gray boxplots show the distribution of results of models with permuted class labels. **(G)** Top: Scatter plot of the absolute value of weights for the classification problem on *stimulus + choice* and on *choice*, utilizing time window of the length $K = [300, 400, 500]$ ms. Weights are collected across recording sessions. Bottom: Correlation coefficient (green) for the data on the top plots Gray box-plots show the distributions of the same results when weights are computed with randomly permuted class labels.

The failure of predicting the variable *choice* from spike counts demonstrates that spike counts of neural populations do not differ between choices "same" and "different" in the absence of the difference in stimuli. The structure of spike counts, however, does differ between the two choices when the choices are both correct, since the variable $stimulus + choice$ is predicted better than chance with the same classification model (Figure 1F, red crosses). From these results, one is tempted to conclude that the activity of neural populations in V1 contains the information about the stimulus class, but not about the choice. An alternative possibility is that the information about the choice cannot be extracted with classical decoding schemes.

We hypothesize that decoding of the choice relies on a partial overlap in the information on the stimuli and the choice. To investigate the overlap in the informational content between *stimulus+choice* and *choice*, we compare decoding weights across the two classification problems. We apply two linear classifiers, one on *stimulus+choice* and one on *choice*, and extract the weights of the classifier in each recording session, getting w_n^{S+C} and w_n^C where n is the neural index. Collecting results across recording sessions, we find a weak but significant positive correlation between the two sets of weights ($r = [0.13, 0.12, 0.13]$ for $K = [300, 400, 500]$ ms, respectively; Figure 1G), indicating the overlap in the informational content of the two classification problems. Another, complementary way to verify the overlap between *stimulus+choice* and *choice* is to consider population vectors. Population vector is a normalized vector of decoding weights for neurons from the same recording session (see methods). We compute population vectors for each classification problem, $\tilde{\mathbf{w}}^{S+C}$ and $\tilde{\mathbf{w}}^C$. The angle between population vectors (Figure 2A) tells about the similarity of the population vectors and, consequently, the similarity of separating hyperplanes in the two classification problems. If population vectors were independent, the angle between them would be orthogonal, $\alpha^{\text{independent}} = \pi/2$, while deviation from orthogonality means that the two vectors are partially overlapping. We find that the average angle between the two population vectors deviates significantly from orthogonality (Figure 2B). The deviation from orthogonality is robust to the length of the time window that we use for decoding ($\alpha = [80.2,\ 80.8,\ 80.0]$ deg for the length of the time window $K = [300,\ 400,\ 500]$ ms, respectively) and is significant in all cases ($p < 0.001$, in all cases, permutation test).

After examining the structure of population spike counts and noticing the similarity between classification problems $stimulus + choice$ and *choice*, our aim is to utilize this similarity to decode the variable *choice* as a time-resolved signal. There are two elements to our decoding model, decoding weights and the population signal, and we assume that the two evolve on very different time scales. Cortical neurons were shown to change their firing rates on the time scale of a couple of milliseconds (Shadlen and Newsome 1998). Rather than with firing rates, the brain might therefore encode the choice dynamically, with a signal that rapidly changes within the trial. Following this idea, we compute the population signal without temporal averaging and keep the time resolution of the recorded spiking signal ($\Delta t = 1$ms). To compute the population signal, spike trains of the neural population are weighted, summed up and convolved with an exponential filter (Figure 2C; see methods), modeling the post-synaptic current at the read-out neuron. Summing of the activity across neurons is justified by the relevance of pooling of neuronal signals for perception (Andrei et al. 2019), and the convolution models the dynamics of the post-synaptic potential. While the population signal is a fast signal that changes on the time scale of a couple of milliseconds, decoding weights change on a much slower time scale. Decoding weights model a quantity proportional to synaptic weights, that change during the training phase of the experiment due to synaptic plasticity. In the brain, synaptic weights change on a time scale of hours or days Abbott and Nelson (2000), much slower than the population signal. We therefore separate the time scales of the two processes and assume that weights are fixed during the computation of the population signal. This is well justified at least in case of V1, where decoding weights with the variable $stimulus + choice$ are relatively stable during the trial and are positively correlated between the first and the second half of the trial, as shown in our previous work (see Koren et al. 2020, Figure 3).

Our decoding procedure consists of the learning and the validation phase, introducing several modifications to the classical decoding scheme in the validation step. In the learning phase, an optimal linear classifier (linear Support Vector Machine) receives samples of population spike counts, where one sample is a vector of z-scored spike counts of simultaneously recorded neurons in a single trial (see methods). The classifier finds the subspace in the high-dimensional neural activity that is optimal for separating conditions "correct match" and "correct

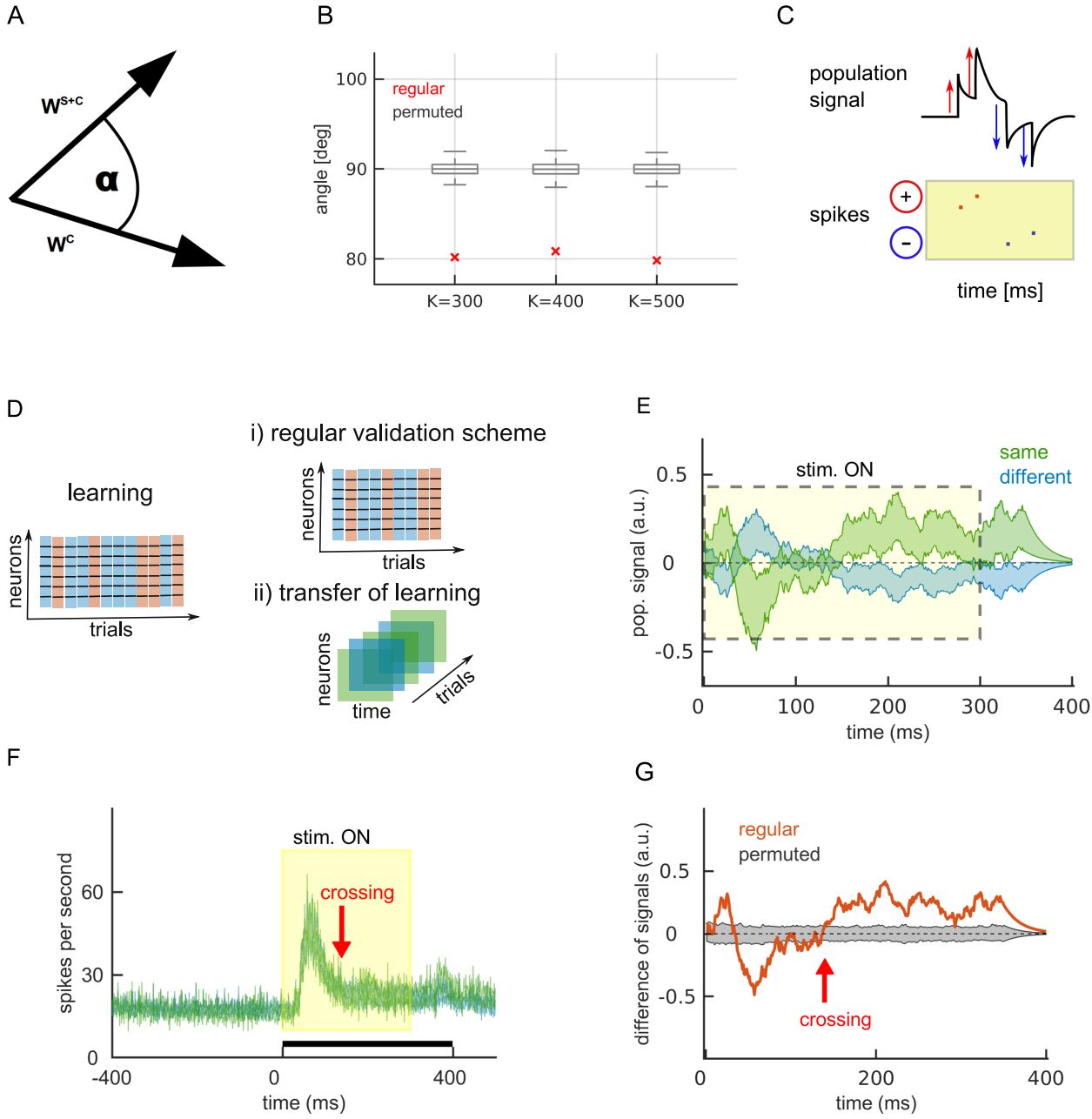


Figure 2. Behavioral choice can be discriminated with generalized learning of neural representation. (A) Schema of the angle between the population vectors in the context of *stimulus + choice* and *choice*. Population vectors are computed in each recording session. **(B)** Average angle between population vectors (red cross) and the distribution of same results for models with random weights (gray box-plot), for the length of the time window $K = 300$ (left), $K = 400$ (middle) and $K = 500$ (right). Parameter: $N_{perm} = 1000$. **(C)** Schema of decoding by weighting spikes. Neurons with positive weights push the signal up and neurons with negative weights pull the signal down. **(D)** Learning scheme. In the learning phase, the classifier learns from samples of parallel spike trains (left). Following the regular validation scheme, the same type of data is used for validating the predictive power of the model on a hold-out set (top right). Here, we test the generalization of learning (bottom right), where the features of the learned classification model are applied on spike trains and on a different classification problem. **(E)** The choice signal for the choice "same" (green) and "different" (blue). We show the mean ± SEM for the variability across sessions. Parameter: $\lambda^{-1} = 20$. **(F)** Population PSTHs for choices "same" (green) and "different" (blue), with the mean ± SEM. The yellow window indicates the presence of the stimulus and the black bar indicates the time window used for analysis. The red arrow marks the time of the zero crossing of the choice signal in **E**. **(G)** Session-averaged difference of the choice signals ("same" - "different") for the regular model (red) and the model with permuted class labels (gray). Parameters: $\lambda^{-1} = 20\ ms$, $N_{perm} = 1000$.

non-match" (variable *stimulus* + *choice*). From the classification model, we compute decoding weights (one number per neuron), that summarize the role of each neuron for the classification model. Within a classical validation scheme, learned structure is then tested on a hold-out set of trials, using the same input statistics and the same categories as the ones used for learning. To decode the population signal corresponding to the choice (the "choice signal"), we deviate from the standard decoding procedure in two ways (Figure 2D). First, we train the model on spike counts, while we validate it on spike trains, due to the separation of time scales justified above. Second, the learning and the validation set also differ in categories that they decode, since the training utilizes the variable *stimulus* + *choice* while the test set is applied on the variable *choice*. The result of our decoding scheme is a low-dimensional representation of the choice, a quantity proportional to the synaptic current at a hypothetical read-out neuron. Note that our results are cross-validated and that the learning and the validation set are non-overlapping (see methods).

With generalized decoding scheme (Figure 2D), the choice "same" vs. "different" can be successfully discriminated during the majority of the trial (Figure 2E). The variable *choice* comprises conditions "correct non-match" and "incorrect non-match". Since non-matching stimuli are rotated for a random angle that changes on each trial, decoding of the choice cannot rely on a particular orientation of the test stimulus. Furthermore, since the variable *choice* does not per se contain the information about the stimulus class (both conditions have non-matching stimuli), we argue that decoded choice signal must truly reflect the information about the choice of the animal. Generalized decoding is necessary for a successful prediction of the choice. When the learning and the validation step both utilize the variable *choice* (instead of learning on *stimulus* + *choice*), resulting signals are barely different from zero and the choice cannot be predicted (Supplementary Figure S1).

Weighting spikes appears to be an essential step for successful discrimination, since population PSTHs corresponding to the choice "same" and "different" are highly overlapping (Figure 2F). To asses the significance of the difference between the choice signals, we compute the session-averaged difference of signals ($\Delta x(t) = x^{\text{same}}(t) - x^{\text{different}}(t)$) and compare the difference with the same result from models with randomly permuted class labels (see methods). Choice signals are significantly different during the majority of the trial (Figure 2G) and depend only weakly on the time scale of the convolution, with longer time scales resulting in a smoother signal (Figure 3A). The discrimination is also robust to the change in the length of the time window that we use for decoding, as the choice can also be discriminated with time windows of 300 and 500 ms (Figure 3B,C). We conclude that prediction of the choice from the spiking activity in V1 is possible with the generalized decoding scheme, and that learning of neural representation in the context of *stimulus* + *choice* successfully generalizes to the representation of the *choice* on a hold-out test set. Finally, we note that during the target time window, the information that is necessary for discrimination of matching from non-matching stimuli is not yet available to the visual system, and consistently, prediction of the choice is not possible (Supplementary Figure S2).

We notice that at 140 ms after the onset of the test stimulus, the read-out of the choice switches sides (Figure 2E), resulting in crossing of the zero line of the difference signal (Figure 2G, red arrow). Since the variable *choice* has a condition with incorrect behavior ("incorrect non-match"), and since incorrect behavior likely reflects an error in the inferred choice Beck et al. (2012), we reason that the switch in decoded choice might reflect the change in representation from the choice "different" to the choice "same" (i.e., from correct to incorrect representation) within the trial. The timing of 140 ms after the onset of the stimulus coincides with the end of transient firing at the stimulus onset (Figure 2F, red arrow), suggesting a role for top-down or lateral inputs, as opposed to bottom-up inputs, for the decoded signal. We also notice that after the "switch", the population signal is roughly stable and persists beyond the offset of the test stimulus (Figures 2E, 3C). The maximum deviation of the difference signal from zero depends on the time window that we use for decoding. As we use the time window of [0,500] ms with respect to the stimulus onset, the strongest deviation from 0 is at 405 ms, more than 100 ms beyond the offset of the test stimulus (at 300 ms; Figure see 3C).

In spite of its simplicity (i.e., linear weighted sum of spikes), the population model has several components that may or may not be necessarily required for discrimination. In order to test the necessity of each component, we remove the information of a specific component and test the effect of such a perturbation on discrimination of the choice. We find that the sign of the weight is necessarily required for discrimination (Figure 3D, left),

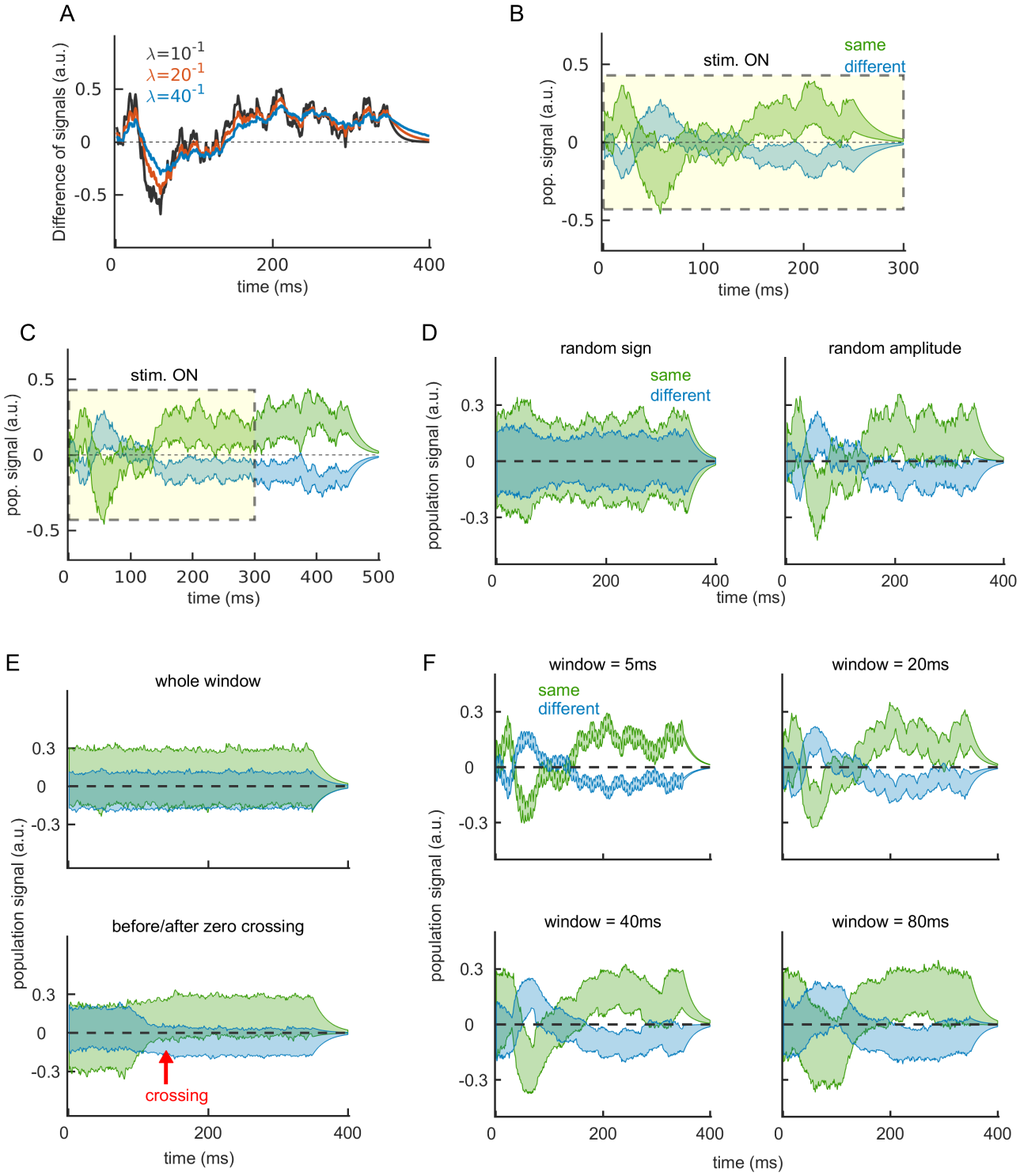


Figure 3. Correct sign of decoding weights is necessary for discrimination. (A) Difference of choice signals for three values of the time constant of the convolution. **(B)** The choice signal during test, utilizing the time window of length $K = 300$ ms. The plot shows the mean ± SEM across recording sessions for the choice "same" (green) and "different" (blue). The time window [0,K] ms with respect to the stimulus onset was used for estimating the population vector as well as for computing the choice signal. **(C)** Same as in **(B)**, but for the length of the time window $K = 500$. **(D)** The choice signals with random sign of weights (top left), random amplitude of weights (top right), permuted spikes across neurons (bottom left) and permuted class labels of spike trains (bottom right). We plot the entire distribution of choice signals in conditions "same" (green) and "different" (blue), where "same" and "different" refer to the animal's choice on non-matching stimuli. **(E)** The choice signals with randomly permuted spike timing across the entire time window (top) and in two time windows, before and after the switch in the read-out of the choice (bottom). Red arrow marks the timing of the crossing (140 ms after the stimulus onset). **(F)** The choice signal with temporal jitter, for different length of the jittering window. Parameters for all plots: $\lambda^{-1} = 20\ ms$, $N_{perm} = 1000$.

while the amplitude of the weight is not necessary, as long as the sign of the weight is correct (Figure 3D, right). However, the sign of the weight is not the only required source of information. When we permute the labels of spike trains (i.e., "same" and "different") in the reconstruction step, the discriminability of the choice is entirely lost (Supplementary Figure S2D). Next, we test the necessity of correct spike timing for discrimination. As we randomly permute the spike timing across the entire time window of 400 ms, we get highly overlapping signals for "same" and "different" (Figure 3E, top). Interestingly, if the spike timing is permuted in the time window only before and after the switch in decoded choice (140 ms after the stimulus onset), the overlap is largely diminished (Figure 3E, bottom), again suggesting different sources of the choice signal before and after this moment. Finally, we ask how does the temporal jitter in short time windows affects the read-out and find that the jitter makes the choice signals noisier, but still allows discrimination (Figure 3F). The read-out of the choice is therefore robust to the perturbation of the spike timing on short time scales.

Noise correlations help the accumulation of the choice signal from spike trains

The sign of the weights is the essential part of the decoding model, since neurons with the opposite sign of the weight have the opposite effect on the classification model, pulling the separation boundary in opposite directions (Figure 4A). We separate the population with respect to the sign of the weight, getting subnetworks of neurons with positive (*plus*) and negative (*minus*) weights. As we compute the choice signal independently with each subnetwork (see methods), we find that the *plus* subnetwork discriminates the choice behavior better than the *minus* subnetwork (Figure 4B). The *minus* subnetwork shows some discriminatory capacity at the very end of the trial (Figure 4B, left), where the response goes in the opposite direction with respect to the signal of the *plus* subnetwork. Measuring the cross-correlation function of the signal of *plus* and *minus* subpopulations in single trials (see methods), we find that the choice signal of the two subnetworks is negatively correlated, consistently for different time constants of the convolution (Figure 4C, top; supplementary table S4.1). This effect is strongly significant (Figure 4C, bottom), and is a signature of competitive dynamics between the two subpopulations. The number of neurons with positive weights (92 neurons) is slightly higher than the number of neurons with negative weights (62 neurons), however, this in unlikely to be important for classification. When decoding *stimulus* + *choice* in a regular fashion, the distribution of weights is centered around zero ($p = 0.235$ for K=400, $p = 0.0842$ for K=500, t-test testing that the mean of the distribution is different from 0; Figure 4D).

We find there are two reasons for the signal of *plus* neurons to be more discriminative. First, weights of *plus* neurons are more informative for classification. Second, patterns of spiking in *plus* neurons are such as to help the accumulation of the population signal. We separate neurons with positive and negative weights and predict the variable *stimulus* + *choice* with a linear classifier, utilizing the information from only one subnetwork at the time (see methods). We find that the predictive power of neurons with positive weights is stronger compared to neurons with negative weights ($p < 10^{-17}$, two-tailed Wilcoxon signed-rank test; Figure 5A). If neurons with the same sign of the weight spike synchronously or are positively correlated on short time scales, this helps the population signal to accumulate (Figure 5B, top) and if neurons with the opposite sign fire in alternation, the population signal remains close to zero (Figure 4B, bottom). Indeed, we find that neurons in the *plus* subnetwork are more strongly synchronized ($p < 0.001$, permutation test; Supplementary Table S5.1) and correlated than neurons within the *minus* subnetwork (Figure 5C).

Besides the sign of the weight, the strength of the weight, $|\tilde{w}_n|$, is important to characterize the impact of the single neuron on the population model, since neurons with strong weight are those that contribute most to the classification. We separate the population of simultaneously recorded units in neurons with strong and weak weights (see methods) and compute the population signal for each subnetwork. As expected, the subnetwork with strong weights is the one that carries the choice signal, while the subnetwork with weak weights performs at chance (Supplementary Figure S3). Moreover, choice signals from the strong and weak subnetworks are positively correlated (Supplementary Table S4.2). As we compute spiking synchrony and correlations within each group and compare results across groups, we find that neurons with strong weight have stronger synchrony

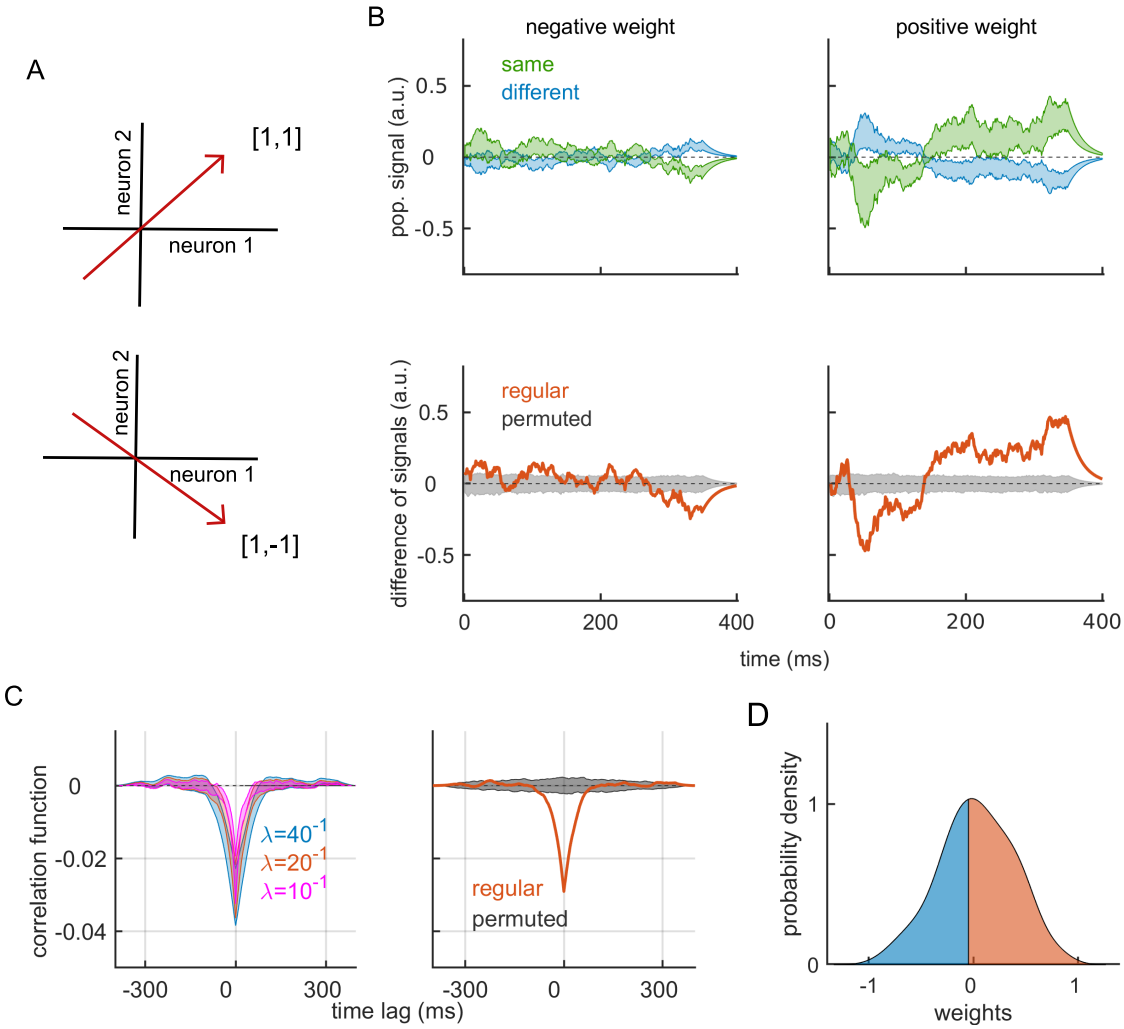


Figure 4. Population signals of neurons with positive and negative weights is negatively correlated. (A) Schema of the effect of the sign of the weight on the orientation of the separation boundary (red arrow). In a toy model with 2 neurons, the upper schema shows the case where the weights of both neurons are positive, and the lower schema illustrates the case with positive and negative weight. **(B)** Top: the choice signal for neurons with negative (left) and positive weight (right). Bottom: Session-averaged difference of signals for the true model (red) and for models with permutation (gray area). **(C)** Left: Correlation function between the choice signal of *plus* and *minus* subnetworks for different values of the time constant of the convolution. Right: Session-averaged correlation function for the regular model (red) and the distribution of results for the model with permutation (gray area). **(D)** Distribution of weights. There are 62 and 92 neurons with negative and positive weight, respectively. Parameters: $\lambda^{-1} = 20$ ms, $N_{perm} = 1000$.

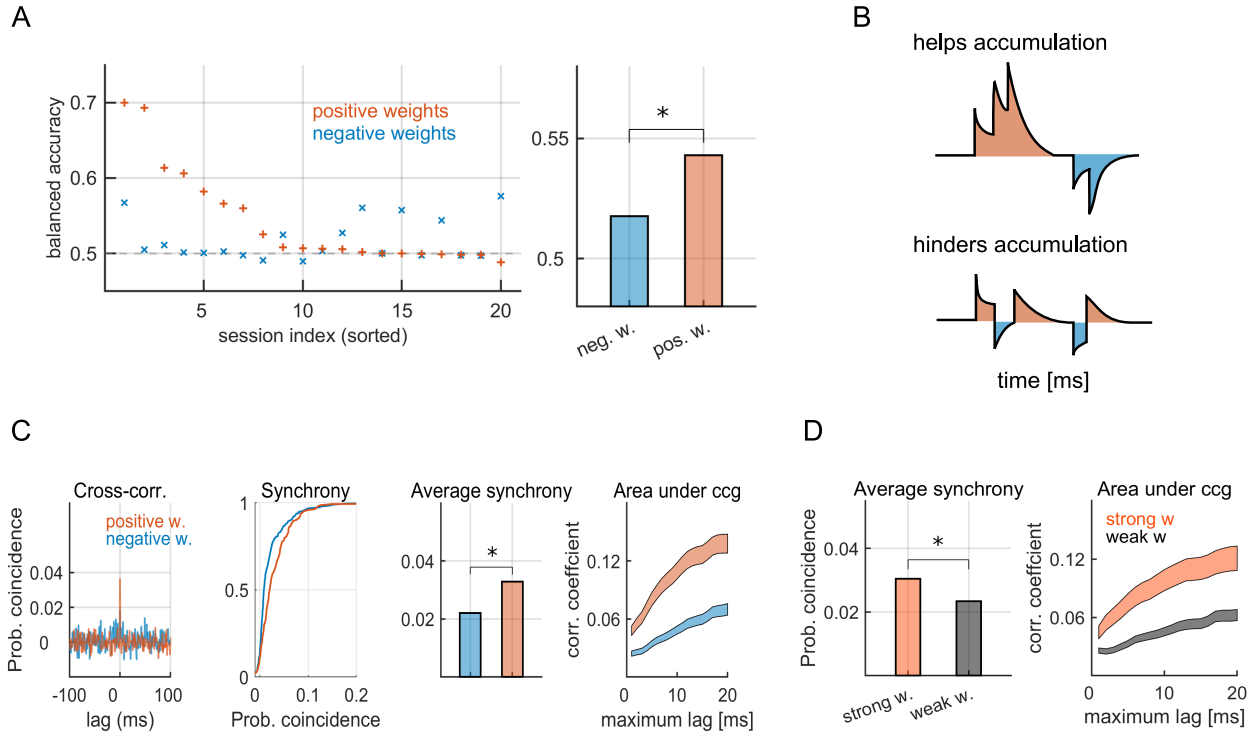


Figure 5. Noise correlations among neurons with positive weights help accumulation of the choice signal. (A) Prediction accuracy of the *stimulus* + *choice* utilizing the information from the *plus* (red) and *minus* (blue) subnetwork. Left plot shows results in sessions, while the right plot shows session-averaged results. The difference in accuracy between *plus* and *minus* subnetworks is tested with t-test on 100 instances of the balanced accuracy with random permutation of the class label (without repetition). **(B)** Schema of spike patterns that either help or hinder the accumulation of the signal. Let's have a pattern of 5 spikes, three spikes of *plus* neurons and 2 spikes of *minus* neurons. The pattern where spikes of neurons with the same sign of the weight are grouped together results in accumulation of the signal (top), while alternative firing of *plus* and *minus* neurons hinders the accumulation of the signal (bottom). **(C)** Synchrony and correlations within the group of *plus* (red) and *minus* (blue) neurons. Left: example cross-correlograms (*ccg*) for a pair of neurons with positive weights (red) and negative weights (blue). Middle left: empirical cumulative distribution functions of the *ccg* at zero lag, measuring pair-wise synchrony. The distribution shows all neuronal pairs, collected across recording sessions (288 and 143 pairs of *plus* and *minus* neurons, respectively). Middle right: average synchrony between pairs of neurons with positive and negative weights. Right: Area under the *ccg* (also known as r^{ccg}), for the increasing length of the summation window. **(D)** Same as **(C)** middle right and right, but for pairwise interactions within the group of neurons with strong (orange, N=147) and weak (dark gray, N=176) weights. The separation in strong and weak weights is done on the classification of *stimulus* + *choice*, while the correlation function is computed for conditions corresponding to the variable *choice*. Parameters: $\lambda^{-1} = 20$ ms, $N_{perm} = 1000$.

and correlation than neurons with weak weights ($p = 0.0140$ permutation test, Figure 5E; Supplementary Table S5.2). This result is an additional argument in favor of the hypothesis that synchrony and correlations on short time scales are useful dynamical features for encoding of the choice in V1.

Activity of bursty neurons and of neurons in the superficial layer is particularly informative about the choice

Besides correlations and synchrony, another way to help the accumulation of the signal might be bursting of single neurons. Bursting helps the accumulation of the signal for the same reason as correlated spiking of neurons with the same sign of decoding weight. We divide neurons into bursty and non-bursty, based on characteristics of the power spectrum of spike trains. The power spectrum of bursty neurons is characterized by a reduced power for middle range frequencies ((Compte et al. 2003), see Figure 6A). We capture this effect by computing the normalized area under the power spectrum, and neuron as bursty if it has the area under the power spectrum smaller than the lower bound of models with permuted spike timing (Figure 6B; see methods). As we compute the choice signal for bursty and non-bursty subnetworks, we find that the bursty subnetwork is largely better at predicting the choice than the non-bursty subnetwork (Figure 6c). Choice signals of bursty and non-bursty subnetworks are positively correlated (Figure 6d; supplementary table S4.3).

Similarly to *plus* neurons, bursty neurons are better at predicting the choice because they have more informative weights and because they are more strongly correlated than non-bursty neurons. Bursty subnetworks predict the variable $stimulus + choice$ better than the non-bursty subnetworks ($p < 10^{-17}$, two-tailed Wilcoxon signed-rank test; Figure 6e). Consistently, the more bursty the neuron, the stronger weight it tends to have. The area under the power spectrum correlates negatively with the strength of the weight, but only for neurons with positive weight ($p = 0.012$ for neurons with positive weights, $p = 0.610$ for neurons with negative weight, permutation test; Figure 6f). In addition, spiking synchrony is stronger between bursty compared to non-bursty neurons ($p < 0.001$, permutation test; Figure 6g; supplementary table S5.3). Bursty neurons also have much stronger noise correlations of spike timing (Figure 6g, right).

At last, we divide neurons with respect to their location along the cortical depth in three layers. Cortical layers in V1 have been shown to importantly differ in their correlation structure (Hansen et al. 2012), which might have direct implications on the population code. Using a method based on the current source density (Koren et al. 2019), we distinguish the superficial layer (supragranular, SG, N=48), the middle layer (granular, G, N=51) and the deep layer of the cortex (infragranular, IG, N=61). As we compute the choice signal in each layer, we find that the superficial layer is by far the most important for discriminating the choice (Figure 7A). The superficial layer also reflects the crossing of the read-out at around 140 ms after the stimulus onset, seen in the overall response. Choice signals are weakly positively correlated across all pairs of layers, consistently for all tested time constants of the convolution (Figure 7B; supplementary table S4.4).

Measuring pair-wise synchrony of spike timing with the cross-correlogram (Figure 7C, left), we find that the synchrony is stronger in the middle and the superficial layer compared to the deep layer ($p < 0.001$ for both SG/IG and G/IG, permutation test), while the difference between the middle and the superficial layer is not significant ($p = 0.573$, Figure 7C, middle plots). The strength of the synchrony in the middle and the superficial layer is therefore comparable (Supplementary table S5.4), however, the superficial layer clearly shows stronger correlation of spike timing (Figure 7C, right), measured as the area under the cross-correlogram. Besides strongest spike-time correlation, the fact that there is excess of bursty neurons and of neurons with positive weights in the superficial layer (Figure 7D) explains why the superficial layer has stronger discriminatory capacity compared to other layers. In the superficial layer, 58 % of neurons have positive weights, while in the middle and deep layers, the percentage of neurons with positive weights is 49 and 51 %, respectively. In the superficial layer, 54 % of neurons are bursty, while in the middle and deep layer, 39 and 36 % of neurons are bursty, respectively.

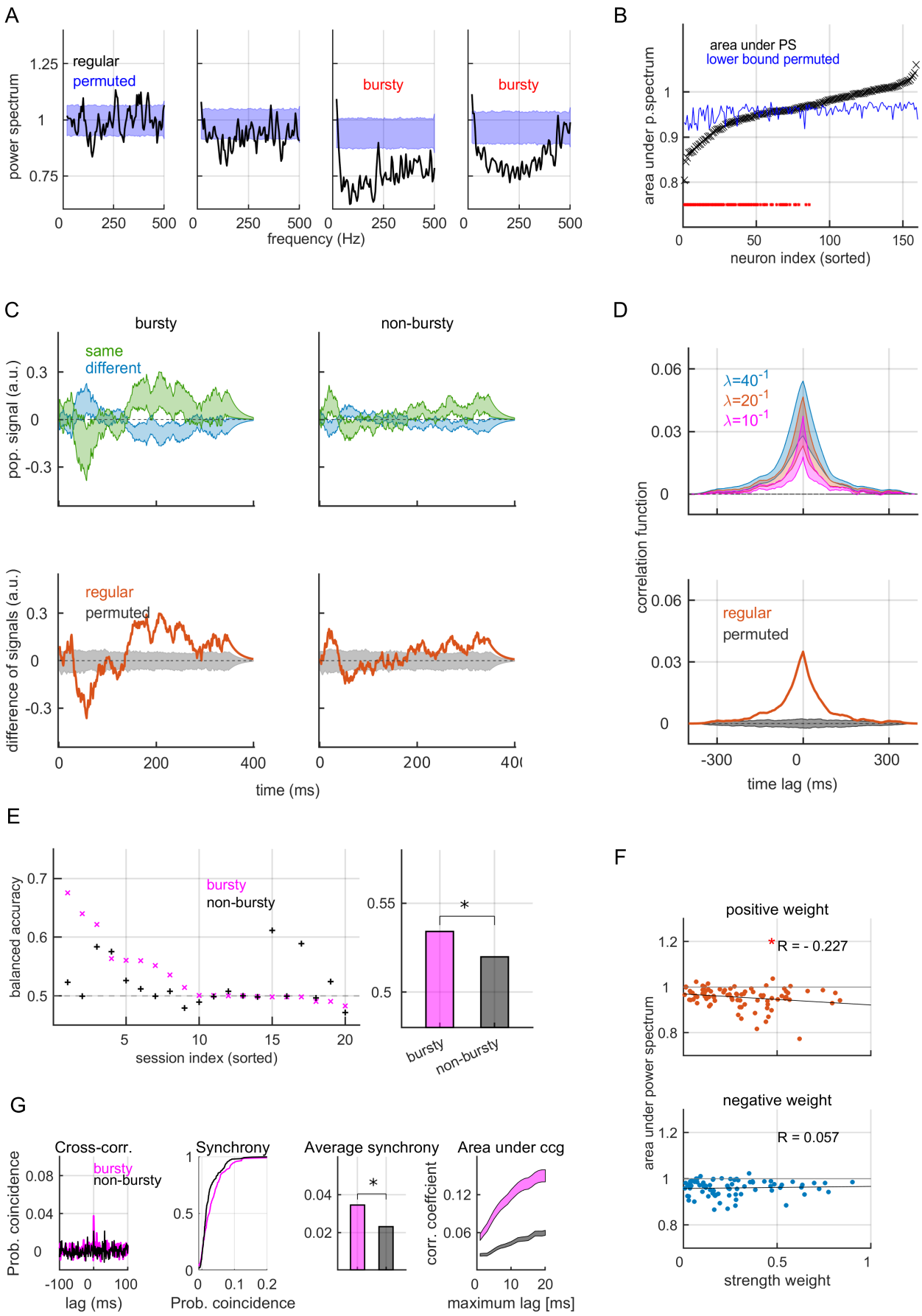


Figure 6. Bursty neurons discriminate better and earlier in the trial. (A) Power spectrum for 4 representative neurons from the same recording session. We show the power spectrum for the true model (black) and the distribution of power spectra for models with randomly permuted spike timing (blue). Bursty neurons are those with reduced power in lower and medium range frequencies (middle right and right).

Figure 6. **(B)** Area under the power spectrum for all neurons, sorted from the lowest to the highest value (black), and the lower bound of the area under the power spectrum for models with permuted spike timing (blue). Red asterisk marks significance (permutation test, $\alpha = 0.05$). There are 68 bursty and 92 non-bursty neurons. **(C)** Top: The choice signal for bursty (left) and non-bursty subnetworks (right), showing the mean $\pm SEM$ in recording sessions. Bottom: Session-averaged difference of signals for the true model (red) and for models with random permutation (gray area). Parameter: $\lambda^{-1} = 20\ ms$. **(D)** Top: Correlation function between the choice signal of bursty and non-bursty subnetworks. We show the mean $\pm$ SEM for the variability across recording sessions. Bottom: Session-averaged correlation function for the regular model (red) and the distribution of results for the model with permutation (gray area). **(E)** Population signal for bursty (magenta) and non-bursty (black) subnetworks. **(F)** Scatter plot of the area under the power spectrum vs. strength of the weight, for neurons with positive weights (top) and neurons with negative weights (bottom). We used trials corresponding to the variable $stimulus + choice$. R marks the Pearson correlation coefficient. **(G)** Same as in Figure 4C, but testing interactions within the group of bursty (magenta) and non-bursty (black) neurons. There are 198 pairs of bursty and 231 pairs of non-bursty neurons. Parameter: $N_{perm} = 1000$.

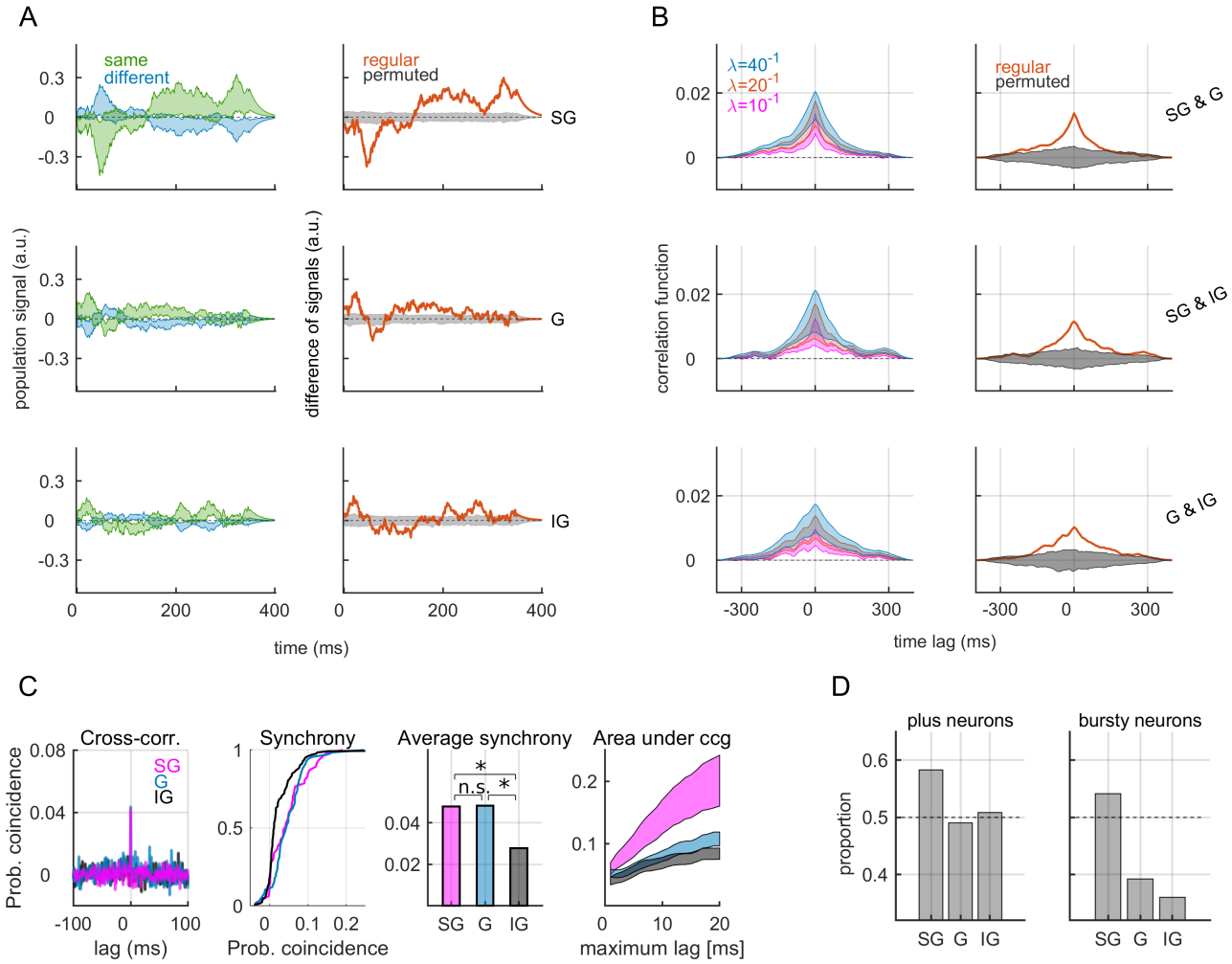


Figure 7. Spiking activity in the superficial layer of the cortex is the most discriminative of the choice behavior. (A) Left: The choice signal in the SG (top), G (middle) and IG layer (bottom), showing the mean ± SEM across recording sessions. Right: Session-averaged difference of choice signals for the true model (red) and the model with permuted class labels (gray). Parameters: $\lambda^{-1} = 20\ ms$, $N_{perm} = 1000$. **(B)** Left: Correlation function of choice signals across pairs of layers. We show the mean ± SEM across recording sessions and plot results for three values of the time constant of the convolution. Right: Session-averaged correlation function for the regular model (red) and for models with permutation (gray area). We use the time constant $\lambda^{-1} = 20\ ms$. **(C)** Same as Figure 5D, but distinguishing three cortical layers. **(D)** The proportion of neurons with positive weight(left) and the proportion of bursty neurons (right) in cortical layers. The proportion is computed relative to the tested group (numbers do not sum to 1 across layers).

Discussion

In this work, we decoded the future behavioral choice of the animal from the spiking activity in the primary visual cortex. We showed that the choice can be decoded as a weighted sum of spikes of neural populations, and that decoding weights need to be learned in correct trials. We argued that the generalization of learning from a mixed variable $stimulus + choice$ to $choice$ is due to the partial overlap in the representation of the stimuli and the choice. Furthermore, we found that bursty neurons and neurons in the superficial layer of the cortex are particularly informative for the task. Our results suggested that bursting of single neurons and correlated spiking of neurons with similar decoding selectivity are dynamical mechanisms that help the formation of the time-resolved choice signal.

Current results raise a question, why is learning in correct trials necessary for the prediction of the choice? The following scenario during the condition "incorrect non-match" is compatible with all our observations: 1) the test stimulus is "non-match" and V1 responds with a spiking pattern corresponding to the stimulus class "non-match" during the first 100 or so milliseconds after the onset of the stimulus, 2) there is an error in transmission of the signal to downstream areas that will result in incorrect choice (Shahidi et al. 2019) 3) due to the error in transmission, a high-level area erroneously signals the stimulus class "match" and the choice "same" and projects this signal back to V1, and 3) as the projection is received by V1, the spiking pattern in V1 switches and signals the stimulus class "match" and the choice "same". This scenario allows the activity in V1 to be predictive of the choice simply through the alignment of the representation about the stimuli and the choice, activated by a top-down projection. Such alignment can be instrumental in the brain, as it reinforces the representation of the upcoming behavioral choice across brain networks.

Another result that merits the discussion is the role of noise correlations of spike timing for the accumulation of the choice signal. Noise correlations of spike counts between neurons with similar selectivity are known to decrease the quantity of information in a rate code (Averbeck et al. 2006), and they do so also in the present dataset (Koren et al. 2020). Recently, the work by (Zylberberg et al. 2017) pointed out that noise correlations among neurons with similar tuning are useful because they improve the robustness of transmission of the rate code downstream. In the present work, we show how structured noise correlations of spike timing can be instrumental in a time-resolved population code. Correlation of spike timing between neurons with similar decoding weights allow the the accumulation of the population signal in time and make the signal rise from the noisy background. Note that if the spiking activity of neurons with the same sign of the weight was entirely uncorrelated, the read-out of such uncorrelated spiking would be a noisy oscillation around zero, precluding discrimination of the choice. Present results therefore suggest a simple but useful mechanism that leverages the structure of correlated spike timing to build the choice signal and help the transmission of the choice signal downstream.

Previous work has suggested that the structure of noise correlations in V1 is primarily determined by a common input (Nienborg et al. 2012) from a top-down source (Bondy et al. 2018). In line with these results, we showed that the spiking synchrony and noise correlations of spike timing are stronger between neurons that are also carriers of the choice signal. As long as top-down projections are specific about the sign of decoding weights of V1 neurons, these projections can at the same time drive the choice signal and create the observed pattern of noise correlations where neurons with similar weights are more strongly correlated. It is sufficient that the top-down input selectively drives neurons with positive weights in trials with decision "same" and neurons with negative weights in trials with decision "different". Precise long range microcircuit-to-microcircuit connections between the prefrontal cortex and V1, recently put in evidence (Ren et al. 2019), are a good candidate for conveying such a feedback signal.

Besides noise correlations of spike timing, we have suggested that bursting of single neurons is a useful dynamical feature for encoding of the choice signal. While bursting might help the accumulation of the choice signal, bursting of single neurons can also increase the reliability of transmission to downstream areas by counteracting the synaptic failure (Lisman 1997), or by driving synchronization in Gamma frequency (Onorato et al. 2020). Bursting neurons observed in our data might correspond to recently reported class of bursting

excitatory neurons, located predominantly in the superficial layers of V1 of the macaque (Onorato et al. 2020). Moreover, consistently with our hypothesis, bursting neurons have been proposed to integrate bottom-up and top-down inputs Larkum (2013); Naud and Sprekeler (2018).

The role of feedforward and feedback inputs for the representation of the choice in sensory areas is a matter of ongoing discussion (Cumming and Nienborg 2016; Nienborg and Cumming 2009; Nienborg et al. 2012; Nienborg and Cumming 2014; Wang 2002), and is also beyond the scope of the present work. Nonetheless, several results of the present analysis allow to discuss about the sources of the choice signal in the present match-to-sample visual task. One of them is the switch in decoded choice at about 140 ms after the stimulus onset. We observed that the difference of signals $\Delta\ x(t) = x^{\text{same}}(t) - x^{\text{different}}(t)$ crosses the zero line at about 140 ms after the stimulus onset (Figure 2E), implying a switch in decoded choice before and after this moment within the trial. The timing of the switch coincides with the qualitative change of the population response from the strong transient at the stimulus onset to tonic firing thereafter (Figure 2F). It is generally accepted that transient elevation of firing rates at the stimulus onset is dominated by the flow of feedforward current to the network, while feedback sources are active during tonic firing that follows (Callaway 2004). The switch in decoded choice therefore raises the possibility of conflicting information from feedforward and feedback sources to the network during trials with incorrect behavior. Note that the timing of the switch in decoded choice is consistent across recording sessions, indicating a precise and reliable timing of the feedback signal across recording sessions. Precise timing of the feedback signal has been shown to be important for performing computations in recurrent networks (Spoerer et al. 2020) and crucial for perception (Silvanto et al. 2005; Hupé et al. 1998). Supposing that in incorrect trials top-down and bottom-up signals are incongruent, we would expect to have congruent top-down and bottom-up signals in correct trials. In our previous work, we have indeed shown that with correct behavior, the variable $stimulus + choice$ is stable across the trial (Koren et al. 2019).

Several of other results from this study suggest that feedback is involved in the processing of choice-related information. If the the feed-forward drive alone would have been decisive for decoding of the choice, we would expect to decode the choice already in the middle layer, which is not supported by present results. Moreover, we have observed that permuting spike timing over the entire time window destroys discriminability of the choice, while limiting the permutation to the time window before and after the switch in decoded choice, discrimination in preserved. This result supports the hypothesis of several sources of the choice signal. Another piece of evidence in favor of the importance of feedback projections is the choice-specific divergence of population signals for neurons with positive and negative weights towards the end of the trial, as well as the negative correlation of population signals between neurons with positive and negative weights. Note that population signals for all other groupings (e.g., bursty/non-bursty, strong/weak weights, layers) were positively correlated. These effects can be explained by a top-down input that selectively drives either neurons with positive of negative weights, depending on the condition.

Recent study in monkey middle temporal area has suggested that the stimulus- and the choice-related information lie in partially distinct subspaces, with a substantial component of the choice-related information in a subspace orthogonal to the stimulus subspace (Zhao et al. 2020). These results are compatible with our findings, since we found that the population vectors corresponding to variables $stimulus + choice$ and $choice$ are only weakly similar, allowing for a large component of neural responses to lie in orthogonal subspaces. In the present data, population vectors deviate for about 10 degrees from orthogonality, and therefore deviate for about 80 degrees from alignment. In (Zhao et al. 2020), authors suggest that the benefit of the partial misalignment of the stimulus and the choice is the co-existence of two representations, with the feedback that targets the "choice" subspace without corrupting the representation in the "stimulus" subspace. Our study, on the other hand, points out the benefits of the partial alignment of subspaces containing the information on the stimulus class and on the choice. In has to be pointed out that while the primary visual cortex, the first stage of cortical processing, is unlikely to have a representation of the choice that is independent from the representation of the stimulus, it might be that in higher cortical areas the representation of the stimulus and of cognitive variables such as the choice might get increasingly more independent. As a hypothesis for the future work, the choice (as well as other cognitive variables) could be represented in multiple sensory areas in parallel, where

such multiplicity of representations would serve the purpose of reverberating and amplifying the choice-related signal within the cortex (Bondy et al. 2018). Such reverberation could implement a form of working memory, by sending the information about a relevant variable across cortical areas, maintaining relevant information within the cortical circuitry.

Acknowledgments

This work was supported by Deutsche Forschungsgemeinschaft (grant number GRK 1589/2).

Author contributions

Conceptualization: V.K. Methodology: V.K., A.A., M.H., V.D., K.O.; Software: V.K.; Validation: V.K.; Formal Analysis: V.K.; Investigation: A.A., M.H.; Resources: V.D., K.O.; Data curation: A.A. and V.K.; Writing-Original Draft: V.K.; Writing-Review and Editing: A.A., V.D., K.O., V.K.; Visualization: V.K.; Supervision: V.D., K.O.; Project Administration: V.K., A.A.; Funding Acquisition: K.O.

Declaration of interests

The authors declare no competing interests.

Methods

Resource availability

Lead contact Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Klaus Obermayer (klaus.obermayer@tu-berlin.de).

Materials availability This study did not generate new unique reagents.

Data and code availability Code is available in a public GitHub repository: https://github.com/VeronikaKoren/transfer learning

Experimental model and subject details

Ethics statement All experiments were conducted in accordance with protocols approved by The Animal Welfare Committee (AWC) and the Institutional Animal Care and Use Committee (IACUC) for McGovern Medical School at The University of Texas Health Science Center at Houston (UTHealth), and met or exceeded the standards proposed by the National Institutes of Health's Guide for the Care and Use of Laboratory Animals.

Animal subjects

The experimental method and animal subject details were described in our previous work (Koren et al. 2019). In the following, we provide a brief summary of the experiment. Two male rhesus macaques (*Macaca mulatta*; M1, 7 years old, 15kg; M2, 11 years old, 13kg) were used in this study. Animals were trained on a delayed match-to-sample task with visual stimuli. As the animal successfully fixated within the fixation area, this automatically triggered the trial. Two naturalistic images appeared consecutively, with a delay period in between. The target and the test stimulus were shown for 300 ms each, with a variable delay period in between (800 to 1000 ms). The identity of stimuli, complex naturalistic images depicting an outdoor scene, changed from one recording session to the other. Within the same session, the target and the test stimuli could be either the same (condition "match") or else the test stimulus was rotated with respect to the target (condition "non-match").

The difference in orientation was such as to keep the performance stable at around 70 percent correct on non-matching stimuli, and ranged between 3 and 10 degrees. Animal subjects were trained to communicate their choice by holding the bar for the choice "different" and releasing the bar for "same". For the trial to be valid, the behavioral response had to be communicated between 200 and 1200 ms after the offset of the test stimulus.

Method details

Multi-unit signal was measured with laminar arrays comprising 16 channels with 100 $\mu-$ meter spacing between adjacent contacts. Electrodes captured neuronal activity across the cortical depth, from superficial to deep layers. The multi-unit signal was spike sorted and we analyzed all cells that responded to the stimulus with at least a 4-fold increase of the firing rate with respect to the baseline. The data was collected in 20 recording sessions, which gave 160 neurons. We analyzed 3 conditions, "correct match" (CM), "correct non-match" (CNM) and "incorrect non-match" (INM), where "correct/incorrect" refers to the behavioral performance and "match/non-match" refers to the stimulus class. In condition "incorrect match", there were not enough trials to perform the analysis. Table 1 reports the summary statistics on the number of trials (biological replicates) in each condition.

Condition	Nb. trials min.	Nb. trials max.	Nb trials average
CM	51	290	118
CNM	42	230	107
INM	21	98	39

Table 1. The minimal, maximal, and average number of trials across recording sessions, for each condition.

Quantification and statistical analysis

The analysis was done with Matlab R2019, Mathworks.

The spike train of a single neuron n in trial j is defined as a binary vector of zeros and ones,

$$o_{n,j}(t_k) = \begin{cases} 1, & \text{if neuron } n \text{ in trial } j \text{ spikes during the } k\text{-th millisecond} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $n = 1, ..., N$ is the neural index, $k = 1, ..., K$, is the time index with step of 1 millisecond, and $j = 1, .., J_1, J_1 + 1, ..., J_2, J_2 + 1, ...J$ is the trial index. Trials were collected in conditions correct match ($j = 1, ..., J_1$), correct non-match ($j = J_1 + 1, ..., J_2$) and incorrect non-match ($j = J_2 + 1, ..., J$).

The coefficient of variation for the neuron n in trial j is defined as follows,

$$CV2_{n,j} = 2\frac{|ISI_{i+1} - ISI_i|}{ISI_{i+1} + ISI_i}$$

where ISI is the inter-spike interval with index $i = 1, ..., N_{int}$. We report trial-averaged results, distinguishing trials from conditions correct non-match (decision "same") and incorrect non-match (decision "different").

$$\begin{aligned} CV2_n^{\text{different}} &= \frac{1}{J_2 - J_1} \sum_{j=J_1+1}^{J_2} CV2_{n,j} \\ CV2_n^{\text{same}} &= \frac{1}{J - J_2} \sum_{j=J_2+1}^{J} CV2_{n,j} \end{aligned} \tag{2}$$

Spike counts were computed in the time window of $[0, K]$ ms with respect to the onset of the stimulus (target or test), $s_{n,j} = \sum_{t_k=1}^{K} o_{n,j}(t_k)$. For the parameter K we used $K = [300, 400, 500]$, as reported in the Results

section. Spike counts were z-scored:

$$\tilde{s}_{n,j} = \frac{s_{n,j} - \langle s_{n,j} \rangle_j}{\sqrt{\mathrm{Var}_j\ (s_{n,j})}}, \tag{3}$$

where $\langle s_{n,j} \rangle_j$ and $\mathrm{Var}_j\ (s_{n,j})$ are the mean and the variance across trials from all conditions.

Prediction accuracy of a linear classifier Z-scored spike counts of N neurons that have beer recorded in parallel are utilized as input to the linear classification model (linear Support Vector Machine, SVM). The prediction accuracy of the classifier is computed as the balanced accuracy on the hold-out test set:

$$BAC = \frac{TP}{2(TP + FN)} + \frac{TN}{2(TN + FP)}, \tag{4}$$

with TP, FN, TN and FP the number of true positive, false negative, true negative and false positive classifications on the test set, respectively. The balanced accuracy is cross-validated $N_{cv} = 100$-times and reported values are averaged across cross-validations. The significance of the balanced accuracy is evaluated with the permutation test, utilizing $N_{perm} = 1000$ random permutations of class labels without repetition (see (Koren et al. 2020)).

When the population is divided in groups (see below), the prediction accuracy for the group of interest is computed by utilizing all simultaneously recorded neurons, while removing the information from neurons that do not belong to the group of interest (see (Koren et al. 2020)). While class labels for the group of interest are correct, they are randomly assigned for the remaining of neurons, and those therefore only contribute chance-level prediction. The advantage of this method is to remove the confounding factor of imbalanced groups.

Learning of structural features from parallel spike counts

To compute the population signal, we first calculate the population vector in each recording session. The population vector is computed as the feature vector of the linear classifier, utilizing parallel spike counts. We then use the population vector to compute the choice signal on parallel spike trains in single trials, without averaging across time.

The computation of the population vector utilizes trials from condition CM and half of the trials from condition CNM (trials with index $j = 1, ..., J_1, J_1 + 1, ..., \frac{J_2 - J_1}{2}$). The computation of the choice signal is computed on a hold-out set, utilizing the remaining half of the trials from condition CNM and all trials from condition INM (trials with index $j = \frac{J_2 - J_1}{2} + 1, ..., J_2, J_2 + 1, ..., J$). The split of trials in training and reconstruction set in condition CNM is cross-validated with Monte-Carlo method, using $N_{cv} = 100$ random splits. The split of trials in training and reconstruction set is non-overlapping, such that no trials that have been used for training appear in the reconstruction set. All reported results are averaged across cross-validations.

Estimation of the population vector In the N-dimensional space of inputs, one sample is the vector of z-scored spike counts of N simultaneously recorded neurons in trial j, $\mathbf{s}_j = [\tilde{s}_{1,j}, \tilde{s}_{2,j}, ..., \tilde{s}_{N,j}]^T$. Linear SVM searches for an $N-1$-dimensional plane (a hyperplane) that optimally separates points in conditions CNM and CM. The hyperplane is defined as follows,

$$H_0 : \mathbf{w}^T \mathbf{s}_j + b = w_1 s_{1,j} + w_2 s_{2,j} + ... + w_N s_{N,j} + b = 0 \tag{5}$$

where $\mathbf{w}$ is the vector of feature weights and b is the offset of the hyperplane from the origin. On each side of H_0, we can define a hyperplane that verifies the following:

$$\begin{aligned} H_1 : \mathbf{w}^T \mathbf{s}_j + b = -1 \\ H_2 : \mathbf{w}^T \mathbf{s}_j + b = 1 \end{aligned} \tag{6}$$

If the problem is linearly separable, all training samples verify the following the inequality,

$$\begin{aligned} \mathbf{w}^T\mathbf{s}_j + b \leq -1 \text{ if } y_j = -1 \\ \mathbf{w}^T\mathbf{s}_j + b \geq 1 \text{ if } y_j = 1 \end{aligned} \tag{7}$$

where $y_j \in \{-1, 1\}$ is the class label ($y_j = -1$ in condition CNM and $y_j = 1$ in condition CM). Training the linear SVM consists in maximizing the number of correctly classified samples and, at the same time, minimizing the distance between H_1 and H_2, which can be expressed with the Lagrangian.

$$L_p = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{j=1}^{\frac{J_2-J_1}{2}} \lambda_j[y_j(\mathbf{w}^T\mathbf{s}_j + b) - 1] \tag{8}$$

The first term on the right hand side ensures the maximal distance between hyperplanes H_1 and H_2, and the second term on the right ensures correct classification. As the derivative of the Lagrangian with respect to $\mathbf{w}$ is set to zero, we get the expression for the vector of weights.

$$\mathbf{w} = \sum_{j=1}^{\frac{J_2-J_1}{2}} \lambda_j y_j \mathbf{s}_j \tag{9}$$

Since $\lambda_j \neq 0$ only for trials that define the margin (points that lie on $H1$ or on $H2$), the weight vector only depends on support vectors,

$$\mathbf{w} = \sum_{q=1}^{Q} \lambda_q y_q \mathbf{v}_q \tag{10}$$

where $q = 1, 2, ..., Q$ are the support vectors, with $Q < \frac{J_2-J_1}{2}$. The weight vector is normalized, $\tilde{\mathbf{w}} = \mathbf{w}/||\mathbf{w}||$, with $||\mathbf{w}|| = (w_1^2 + ... + w_N^2)^{\frac{1}{2}}$. We refer to the vector $\tilde{\mathbf{w}} = [\tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_n, ..., \tilde{w}_N]^T$ as the population vector and to the $n-$th entry of $\tilde{\mathbf{w}}$, $\tilde{w}_n$, as the decoding weight of the neuron n.

Representation of the choice signal as a read-out of parallel spike trains

In the following, we describe the reconstruction part. For a similar method, see (Koren et al. 2019).

Consider the vector of spike trains of N simultaneously recorded neurons.

$$\mathbf{o}_j(t_k) = [o_{1,j}(t_k), o_{2,j}(t_k)..., o_{N,j}(t_k)]^T \tag{11}$$

The choice signal is a projection of the vector of spike trains on the vector of weights,

$$x_j(t_k) = F(\tilde{\mathbf{w}}^T\mathbf{o}_j(t_k)) - \langle F(\tilde{\mathbf{w}}^T\mathbf{o}_j(t_k))\rangle_j \tag{12}$$

where $\langle F(\tilde{\mathbf{w}}^T\mathbf{o}_j(t_k))\rangle_j$ is the trial average, utilizing all trials of the reconstruction set. By subtracting the trial average, we compute the deviation of the signal from the mean. We argue that the deviation of the signal from its mean, rather than the absolute value of the signal, might be biologically relevant. As the transfer function, $F(y(t_k))$, we use a convolution with an exponential kernel,

$$F(y) = \sum_{\tau \in T} y(t_k - \tau)u(\tau) \tag{13}$$

with $u(\tau) = exp(-\lambda\tau),\ \tau \in T$, with support $T = \{-100, ..., 100\}$ ms. Convolution with an exponential kernel models the causal effect of the presynaptic spike on the neural membrane of the read-out neuron. Notice that $x_j(t_k)$ is a time-resolved, low-dimensional representation of parallel spike trains in single trials.

To test the discriminability of the choice, we average the choice signal across trials, distinguishing conditions

CNM (decision "different") and INM (decision "same"),

$$x^{\text{same}}(t_k) = \frac{2}{J_2 - J_1} \sum_{j=J'}^{J_2} x_j(t_k)$$
$$x^{\text{different}}(t_k) = \frac{1}{J - J_2} \sum_{j=J_2+1}^{J} x_j(t_k) \tag{14}$$

where $J' = \lceil \frac{1}{2}(J_1 + J_2) \rceil + 1$, and $\lceil z \rceil$ is the ceiling function . The significance of the difference between choice signals for the decision "same" and "different" is evaluated with the permutation test. We compute the difference of the choice signals in every recording session,

$$\Delta x(t_k) = x^{\text{same}}(t_k) - x^{\text{different}}(t_k) \tag{15}$$

and average across sessions. We then rank $\Delta x(t_k)$ among $\Delta x_p^{perm}(t_k)$, where the latter has been computed with random weights, and the class labels in the validation set have been randomly permuted. Random weights were drawn from the uniform distribution with the same range as the regular weights. The permutation procedure is repeated N_{nperm}-times and gives a distribution of results for each time step. When the result of the true model, $\Delta x(t_k)$, appears outside of the distribution of results of the null model, $\Delta x_p^{perm}(t_k)$ for $p = 1, 2, ..., N_{perm}$, we consider that signals $x^{\text{same}}(t_k)$ and $x^{\text{diff}}(t_k)$ have been successfully discriminated.

Criteria for division into subpopulations

Sign of the weight We separate the population with respect to the sign of entries of the population vector, distinguishing neurons with positive weight ($\tilde{w}_n > 0$, *plus* neurons) and negative weight ($\tilde{w}_n < 0$, *minus* neurons). The sign of the weight is determined with respect to the variable *stimulus* + *choice* (utilizing conditions correct match and correct non-match). To determine the sign as accurately as possible, we compute the population vector utilizing all available trials from conditions correct match and correct non-match.

Strength of the weight We distinguish neurons with strong and weak weights by ranking the absolute value of the weight, $|\tilde{w}_n|$, among the distribution of weights of models with permuted class labels. The strength of the weight of a single neuron n is a scalar, and the same result for models with permuted class labels is a distribution of N_{perm} values, where N_{perm} is the number of random permutations of class labels. If the strength of the weight of the neuron n is ranked within the first 25 % of weights from the null model, we assume that the neuron n has strong weight, and assume it has weak weight otherwise. Similarly as with the sign of the weight, the strength of the weight is computed using conditions correct match and correct non-match and utilizing all trials from these conditions.

Burstiness We distinguish bursty and non-bursty neurons utilizing a criterion based on the power spectrum of spike trains. It has been shown that bursty neurons have decreased power spectrum in low and middle frequency ranges (Compte et al. 2003). We compute the power spectrum of spike trains for every single neuron, using multiplication of the spike train with Slepian tapers to increase the reliability of the estimation (Pesaran et al. 2002). We use 5 Slepian tapers and the time window of $K = 400$ ms. Power spectra are normalized with neuron's firing rate. We then compute the normalized sum of the power spectrum for frequencies between 10 and 200 Hz, discarding frequencies below 10 Hz due to the short time window K that prevents reliable estimation at low frequencies. As a reference, homogeneous Poisson process has a constant power spectrum of 1, yielding a normalized sum under the power spectrum of 1, while the sum under the power spectrum of a bursty neuron is typically lower than 1.

Significance of the result is estimated with the permutation test, comparing the sum under the power spectrum of the regular model with the distribution of same results for the null model. The null model is computed by following the same procedure as for the regular model, but randomly permuting, without repetition,

the time index of spike trains. We iterate the null model $N_{perm}-$ times and rank the result of the regular model among the distribution of the same results given by null models. We assume that the neuron n is bursty if the normalized sum under its power spectrum is significantly below the same result of models with permutation ($\alpha = 0.05$). If the neuron does not fulfill this criterion, it is assumed to be non-bursty.

Cortical layers We distinguish three cortical layers, the superficial (supragranular, SG), the middle (granular, G) and the deep layer (infragranular, IG) (Hansen et al. 2012). The method for determining cortical layers utilizes the covariance matrix of the current source density and has been published previously (Figure 6 and methods in (Koren et al. 2019)).

The choice signal in subpopulations

Sign-specific choice signal is computed by removing the information from the spike train of neurons with the opposite sign. This method ensures that the imbalance in the number of *plus* and *minus* neurons does not influence the population signal. We remove the information from one of the groups by randomly permuting class labels "same" and "different" in the reconstruction step. When the class label in the reconstruction step is permuted, the association between the weight and the class label of the spike train is random, resulting in a signal that is close to zero at all times. As an example, the choice signal of *plus* neurons is computed with the spike train $\mathbf{o}_{j,p}^{perm,+}(t_k)$, where the label of the spike train is correct for *plus* neurons, and random (i.e., correct or incorrect with equal probability) for *minus* neurons.

$$x_{j,p}^{+}(t_k) = F(\tilde{\mathbf{w}}^T \mathbf{o}_j^{perm,+}(t_k)) - \langle F(\tilde{\mathbf{w}}^T \mathbf{o}_j^{perm,+}(t_k))\rangle_j \tag{16}$$

Similarly, the signal of *minus* neurons is computed by utilizing the spike train $\mathbf{o}_{j,p}^{perm,-}(t_k)$, where the label of the spike train is correct for *minus* neurons, and random for *plus* neurons.

$$x_{j,p}^{-}(t_k) = F(\tilde{\mathbf{w}}^T \mathbf{o}_j^{perm,-}(t_k)) - \langle F(\tilde{\mathbf{w}}^T \mathbf{o}_j^{perm,-}(t_k))\rangle_j \tag{17}$$

Random permutation is repeated $N_{perm}-$times, with $p = 1, 2, ..., N_{perm}$ random permutations, without repetition, of the order of trials. Next, we average each of the signals across trials, distinguishing conditions "same" and "different".

$$\begin{aligned} x_p^{+,\text{same}}(t_k) &= \frac{2}{J_2 - J_1} \sum_{j=J'}^{J_2} x_{j,p}^{+}(t_k) \\ x_p^{+,\text{different}}(t_k) &= \frac{1}{J - J_2} \sum_{j=J_2+1}^{J} x_{j,p}^{+}(t_k) \end{aligned} \tag{18}$$

Same follows for *minus* neurons. The signal is then averaged across permutations, getting $x^{+,\text{same}}(t_k)$ and $x^{+,\text{different}}(t_k)$ as the signal for *plus* subnetwork and $x^{-,\text{same}}(t_k)$ and $x^{-,\text{different}}(t_k)$ as the signal for the *minus* subnetwork. The significance is evaluated with the permutation test. The test statistic is the sign-specific difference of signals in conditions "same" and "different".

$$\begin{aligned} \Delta x^{+}(t_k) &= x^{+,\text{same}}(t_k) - x^{+,\text{different}}(t_k) \\ \Delta x^{-}(t_k) &= x^{-,\text{same}}(t_k) - x^{-,\text{different}}(t_k) \end{aligned} \tag{19}$$

The null model is computed with the random permutation of class labels 1) when training the classification model and 2) in the reconstruction step. In addition, we use a random assignment to the class of *plus* and *minus* neurons by randomly permuting neural indexes.

The same methods is used to compute the choice signal for subnetworks of neurons with strong and weak weights, for bursty and non-bursty subnetworks and in cortical layers.

Correlation function between the choice signals

The correlation function between the choice signals of *plus* and *minus* subnetworks in trial j and for the permutation instance p is defined as follows:

$$R^{+-}_{j,p}(\tau) = \begin{cases} \sum_{k=0}^{K-\tau-1} x^{+}_{j,p}(t_k) x^{-}_{j,p}(t_k+\tau) & \text{for } \tau \geq 0 \\ \\ R^{-+}_{j}(-\tau) & \text{for } \tau < 0 \end{cases} \tag{20}$$

with time lag $\tau = 1, 2, ..., 2K-1$. The correlation function is normalized with autocorrelation functions at zero time lag,

$$\tilde{R}^{+-}_{j,p}(\tau) = \frac{R^{+-}_{j,p}(\tau)}{R^{--}_{j,p}(0) R^{++}_{j,p}(0)} \tag{21}$$

where R^{++} (R^{--}) is the autocorrelation function for *plus* (*minus*) neurons.

$$R^{++}_{j,p}(\tau) = \begin{cases} \sum_{k=0}^{K-\tau-1} x^{+}_{j,p}(t_k) x^{+}_{j,p}(t_k+\tau), & \text{for } \tau \geq 0 \\ \\ R^{++}_{j,p}(-\tau) & \text{for } \tau < 0 \end{cases} \tag{22}$$

The correlation function is computed in single trials and then averaged across trials and across permutations. Since there was no difference in the correlation across conditions, we used all trials from the reconstruction set (conditions CNM and INM, trials with index $j = J_1 + 1, ..., J$).

$$\tilde{R}^{+-}(\tau) = \frac{1}{N_{perm}(J - J_1)} \sum_{p=1}^{N_{perm}} \sum_{j=J_1+1}^{J} \tilde{R}^{+-}_{j,p}(\tau) \tag{23}$$

The significance of the correlation function is estimated with the permutation test. We compute the choice signal with random weights and random class label of spike trains. In addition, we use a random assignment to the group of *plus* and *minus* neurons as we compute the correlation function. The same method is used to compute the correlation function between subnetworks with strong and weak weights, and between bursty and non-bursty subnetworks.

Correlation function of choice signals in cortical layers

Similarly, we compute the cross-correlation of choice signals between pairs of cortical layers. The correlation function is computed for the choice signals from two cortical layers,

$$R^{c_1 c_2}_{j,p}(\tau) = \begin{cases} \sum_{k=0}^{K-\tau-1} x^{c_1}_{j,p}(t_k) x^{c_2}_{j,p}(t_k+\tau), & \text{for } \tau \geq 0 \\ \\ R^{c_2 c_1}_{j,p}(-\tau), & \text{for } \tau < 0 \end{cases} \tag{24}$$

with $(c_1, c_2) \in \{(SG, G), (SG, IG), (G, IG)\}$. The rest of the procedure is the same as for *plus* and *minus* neurons. The significance of results is evaluated with the permutation test, where signals are computed with random weights and random class of the spike train. During the computation of the correlation function, the null model in addition uses random assignment to one of the three cortical layers.

Comparison of population vectors across informational contexts of *stimulus+choice* and *choice*

Similarity of population vectors We compare the population vectors in the context of *stimulus+choice* and in the context of *choice*. We train the linear SVM and compute the population vector in each of the two contexts, getting two vectors, $\tilde{\mathbf{w}}^{S+C}$ and $\tilde{\mathbf{w}}^{C}$, in each recording session. Decoding model in the context of *stimulus+choice* utilizes trials from conditions CM and CNM, while decoding model in the context of *choice* utilizes trials from conditions CNM and INM. The number of trials is imbalanced across conditions (namely, there are less trials in condition INM compared to CM), and such imbalance can affect the population vector. We balance the number of trials with the bootstrap method. In each recording session, we find the number of trials of the condition with most trials, and randomly sample, with repetition, the same number of trials from the two other distributions. All reported results are averaged across bootstraps.

We measure the similarity of the two population vectors by computing the angle between them,

$$\alpha = \arccos\left(\tilde{\mathbf{w}}^{S+C} \cdot \tilde{\mathbf{w}}^{C}\right) \tag{25}$$

where $(\cdot)$ is the dot product between the two vectors. Notice that, since vectors are normalized, we have that $||\tilde{\mathbf{w}}^{S+C}|| = ||\tilde{\mathbf{w}}^{C}|| = 1$. If the vectors $\tilde{\mathbf{w}}^{S+C}$ and $\tilde{\mathbf{w}}^{C}$ are similar, they point in similar direction, and the angle between them is small. If, conversely, the two vectors are pointing in random directions between 0 and π, the angle between them is, on average, orthogonal (the average is across bootstrapped samples). The significance of the angle is evaluated with the permutation test, using the test statistics of the angle, averaged across recording sessions. To construct the null model, we draw random vectors from the uniform distribution that have the same range as the true population vectors, and compute the angle between the two random vectors, $\alpha_p = \arccos\left(\tilde{\mathbf{w}}_p^{S+C} \cdot \tilde{\mathbf{w}}_p^{C}\right),\ p = 1, ..., N_{perm}$. The p-value is computed by ranking the angle of the true model among the distribution of angles of the model with random weights. The test is significant if p-val $< {}^{0.05}/_{N_{test}}$, where the division with the number of tests implements the Bonferroni correction for multiple testing.

Distance of decoding weights While the similarity of population vectors is a population-wise measure, we are also interested in understanding how different are weights of single neurons in the context of *stimulus+choice* and in the context of *choice*. For each single neuron, we compute the distance between its decoding weight in the two contexts.

$$d_n^{S+C,C} = |\tilde{w}_n^{S+C} - \tilde{w}_n^{C}| \tag{26}$$

Results are collected across recording sessions. We then split neurons in groups according to a specific criterion (sign of the weight, strength of the weight, burstiness, layers) and test the difference between groups with a two-tailed t-test.

Acknowledgements

This work was supported by Deutsche Forschungsgemeinschaft, grant GRK 1589/2.

Author contributions

Conceptualization: V.K. Methodology: V.K., A.A., M.H., V.D., K.O.; Software: V.K.; Validation: V.K.; Formal Analysis: V.K.; Investigation: A.A., M.H.; Resources: V.D., K.O.; Data curation: A.A. and V.K.; Writing-Original Draft: V.K.; Writing-Review and Editing: A.A., V.D., K.O.; Visualization: V.K.; Supervision: V.D., K.O.; Project Administration: V.K., A.A.; Funding Acquisition: K.O.

Declaration of interest

The authors declare no competing interests.

References

Abbott, L. F. and Nelson, S. B. (2000). Synaptic plasticity: taming the beast. *Nature neuroscience*, 3(11):1178–1183.

Andrei, A. R., Pojoga, S., Janz, R., and Dragoi, V. (2019). Integration of cortical population signals for visual perception. *Nature communications*, 10(1):1–13.

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nature reviews neuroscience*, 7(5):358–366.

Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., and Pouget, A. (2012). Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron*, 74(1):30–39.

Bernardi, S., Benna, M. K., Rigotti, M., Munuera, J., Fusi, S., and Salzman, C. D. (2020). The geometry of abstraction in the hippocampus and prefrontal cortex. *Cell*, 183(4):954–967.

Blainey, P., Krzywinski, M., and Altman, N. (2014). Points of significance: replication. *Nature methods*, 11(9):879.

Bondy, A. G., Haefner, R. M., and Cumming, B. G. (2018). Feedback determines the structure of correlated variability in primary visual cortex. *Nature neuroscience*, 21(4):598–606.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., Movshon, J. A., et al. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque mt. *Visual neuroscience*, 13(1):87–100.

Callaway, E. M. (2004). Feedforward, feedback and inhibitory connections in primate visual cortex. *Neural Networks*, 17(5-6):625–632.

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., and Shenoy, K. V. (2012). Neural population dynamics during reaching. *Nature*, 487(7405):51–56.

Compte, A., Constantinidis, C., Tegnér, J., Raghavachari, S., Chafee, M. V., Goldman-Rakic, P. S., and Wang, X.-J. (2003). Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *Journal of neurophysiology*, 90(5):3441–3454.

Cumming, B. G. and Nienborg, H. (2016). Feedforward and feedback sources of choice probability in neural population responses. *Current opinion in neurobiology*, 37:126–132.

Gao, P. and Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Current opinion in neurobiology*, 32:148–155.

Hansen, B. J., Chelaru, M. I., and Dragoi, V. (2012). Correlated variability in laminar cortical circuits. *Neuron*, 76(3):590–602.

Harris, K. D., Bartho, P., Chadderton, P., Curto, C., de la Rocha, J., Hollender, L., Itskov, V., Luczak, A., Marguet, S. L., Renart, A., et al. (2011). How do neurons work together? lessons from auditory cortex. *Hearing research*, 271(1-2):37–53.

Hupé, J., James, A., Payne, B., Lomber, S., Girard, P., and Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by v1, v2 and v3 neurons. *Nature*, 394(6695):784–787.

Juan, C.-H. and Walsh, V. (2003). Feedback to v1: a reverse hierarchy in vision. *Experimental brain research*, 150(2):259–263.

Koren, V., Andrei, A. R., Hu, M., Dragoi, V., and Obermayer, K. (2019). Reading-out task variables as a low-dimensional reconstruction of neural spike trains in single trials. *PloS one*, 14(10):e0222649.

Koren, V., Andrei, A. R., Hu, M., Dragoi, V., and Obermayer, K. (2020). Pairwise synchrony and correlations depend on the structure of the population code in visual cortex. *Cell Reports*, 33(6):108367.

Koren, V. and Denève, S. (2017). Computational account of spontaneous activity as a signature of predictive coding. *PLoS computational biology*, 13(1):e1005355.

Larkum, M. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends in neurosciences*, 36(3):141–151.

Lisman, J. E. (1997). Bursts as a unit of neural information: making unreliable synapses reliable. *Trends in neurosciences*, 20(1):38–43.

Logothetis, N. K. and Schall, J. D. (1989). Neuronal correlates of subjective visual perception. *Science*, 245(4919):761–763.

Luczak, A., Bartho, P., and Harris, K. D. (2013). Gating of sensory input by spontaneous cortical activity. *Journal of Neuroscience*, 33(4):1684–1695.

Muckli, L., De Martino, F., Vizioli, L., Petro, L. S., Smith, F. W., Ugurbil, K., Goebel, R., and Yacoub, E. (2015). Contextual feedback to superficial layers of v1. *Current Biology*, 25(20):2690–2695.

Naud, R. and Sprekeler, H. (2018). Sparse bursts optimize information transmission in a multiplexed neural code. *Proceedings of the National Academy of Sciences*, 115(27):E6329–E6338.

Ni, A. M., Ruff, D. A., Alberts, J. J., Symmonds, J., and Cohen, M. R. (2018). Learning and attention reveal a general relationship between population activity and behavior. *Science*, 359(6374):463–465.

Nienborg, H. and Cumming, B. G. (2006). Macaque v2 neurons, but not v1 neurons, show choice-related activity. *Journal of Neuroscience*, 26(37):9567–9578.

Nienborg, H. and Cumming, B. G. (2009). Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature*, 459(7243):89–92.

Nienborg, H. and Cumming, B. G. (2014). Decision-related activity in sensory neurons may depend on the columnar architecture of cerebral cortex. *Journal of Neuroscience*, 34(10):3579–3585.

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

Onorato, I., Neuenschwander, S., Hoy, J., Lima, B., Rocha, K.-S., Broggini, A. C., Uran, C., Spyropoulos, G., Klon-Lipok, J., Womelsdorf, T., et al. (2020). A distinct class of bursting neurons with strong gamma synchronization and stimulus selectivity in monkey v1. *Neuron*, 105(1):180–197.

Panzeri, S., Harvey, C. D., Piasini, E., Latham, P. E., and Fellin, T. (2017). Cracking the neural code for sensory perception by combining statistics, intervention, and behavior. *Neuron*, 93(3):491–507.

Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., and Andersen, R. A. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nature neuroscience*, 5(8):805–811.

Ren, S.-Q., Li, Z., Lin, S., Bergami, M., and Shi, S.-H. (2019). Precise long-range microcircuit-to-microcircuit communication connects the frontal and sensory cortices in the mammalian brain. *Neuron*, 104(2):385–401.

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451):585–590.

Roitman, J. D. and Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of neuroscience*, 22(21):9475–9489.

Saxena, S. and Cunningham, J. P. (2019). Towards the neural population doctrine. *Current opinion in neurobiology*, 55:103–111.

Shadlen, M. N., Britten, K. H., Newsome, W. T., and Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *Journal of Neuroscience*, 16(4):1486–1510.

Shadlen, M. N. and Newsome, W. T. (1996). Motion perception: seeing and deciding. *Proceedings of the national academy of sciences*, 93(2):628–633.

Shadlen, M. N. and Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of neuroscience*, 18(10):3870–3896.

Shahidi, N., Andrei, A. R., Hu, M., and Dragoi, V. (2019). High-order coordination of cortical spiking activity modulates perceptual accuracy. *Nature neuroscience*, 22(7):1148–1158.

Silvanto, J., Cowey, A., Lavie, N., and Walsh, V. (2005). Striate cortex (v1) activity gates awareness of motion. *Nature neuroscience*, 8(2):143–144.

Spoerer, C. J., Kietzmann, T. C., Mehrer, J., Charest, I., and Kriegeskorte, N. (2020). Recurrent neural networks can explain flexible trading of speed and accuracy in biological vision. *PLoS computational biology*, 16(10):e1008215.

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature*, 571(7765):361–365.

Tkačik, G., Marre, O., Amodei, D., Schneidman, E., Bialek, W., and Berry II, M. J. (2014). Searching for collective behavior in a large network of sensory neurons. *PLoS Comput Biol*, 10(1):e1003408.

Uka, T. and DeAngelis, G. C. (2004). Contribution of area mt to stereoscopic depth perception: choice-related response modulations reflect task strategy. *Neuron*, 42(2):297–310.

Wang, X.-J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, 36(5):955–968.

Zhao, Y., Yates, J. L., Levi, A. J., Huk, A. C., and Park, I. M. (2020). Stimulus-choice (mis) alignment in primate area mt. *PLoS computational biology*, 16(5):e1007614.

Zipser, K., Lamme, V. A., and Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *Journal of Neuroscience*, 16(22):7376–7389.

Zylberberg, J., Pouget, A., Latham, P. E., and Shea-Brown, E. (2017). Robust information propagation through noisy neural circuits. *PLoS computational biology*, 13(4):e1005497.