

# Quantifying neuronal selectivity: an information-theoretic approach to calcium imaging data analysis

Nikita Pospelov<sup>1</sup>, Viktor Plusnin<sup>1</sup>, Olga Rogozhnikova<sup>1</sup>, Vladimir Sotskov<sup>2</sup>, Ksenia Toropova<sup>1</sup>, Olga Ivashkina<sup>1</sup>, Vladik Avetisov<sup>3</sup> and Konstantin Anokhin<sup>1</sup>

<sup>1</sup>Institute for Advanced Brain Studies of the Moscow State University, Moscow, Russia

<sup>2</sup> Paris, France ([affiliation](#))

<sup>3</sup>([affiliation](#))

## Аннотация

Brain neurons reflect information about the relevant properties of the environment through their activity. During the individual experience of an animal, neurons become selective for certain aspects of the outside world. Understanding the mechanisms behind the formation, reorganization, and overlap of these selectivity patterns is crucial for understanding brain function.

Recent advances in calcium imaging have made it possible to study these processes in living animals by monitoring multiple neurons' activity during specific tasks or free behavior. This has created a need for tools that can accurately and effectively analyze data to find relationships between neuronal activity, environmental variables, and behavioral characteristics. Here we present an open-source tool based on the calculation of mutual information from time series of neural activity and behavior: INTENSE. Using an information-theoretical approach, INTENSE can identify nonlinear relationships between variables of interest and can work directly with calcium fluorescence signals, while accurately controlling for spurious neuronal selectivities via bootstrap analysis.

Our approach shows promising results on synthetic data, being able to detect neuronal selectivity even in the presence of high noise. We also test INTENSE using the example of mice exploring a new environment while simultaneously recording the activity of hippocampal neurons using a miniscope. We show that there are multiple neurons that are selective for various aspects of environment or animal behavior, including place,

locomotion, object and head direction cells, as well as neurons that reflect several different properties in their activity. We also propose an information-theoretic approach for 'disentangling' such multiple selectivities, which can provide additional insights into neural coding.

## 1 Introduction

### 1.1 Neuronal selectivity

During experience, brain neurons begin to selectively respond to certain elements of the environment. These single neuron selectivities can be simple (as shown in the pioneering work of Hubel and Wiesel, who discovered the selectivity of neurons in the primary visual cortex for simple patterns [1]), or they can depend on complex environmental patterns and even abstract concepts, such as "face neurons" in macaques [2, 3] and humans [4].

Of particular interest for the analysis of cognitive specializations of neurons is their co-relation with environmental variables and animal behavior.

Neuronal selectivity to complex vocalization features emerges in the superficial layers of primary auditory cortex [5] distinct perceptual features of odors [6] интеграция модальностей [7]

Due to the rapid development of large language models (LLMs) in recent years, research on their interpretability has become increasingly important. This involves understanding the role of individual elements in the model's operation. Experiments have shown that many of the selectivity properties found in biological neurons can be replicated in artificial neural networks. For example,

---

Interestingly, just like in living brains, ANN neurons can exhibit selectivity for multiple stimuli [8] Abstract representations emerge naturally in neural networks trained to perform multiple tasks [9]

### 1.2 Information theory and the neural code

Information theory, which originated with Shannon's seminal work [10], has a long history of application in neuroscience. Examples of successful applications of information-theoretic methods include the analysis of both functional [11, 12] and effective [13] large-scale brain connectivity, electroencephalography (EEG) [14, 15], and reconstruction of connectivity of simulated [16] and real [17] neuronal networks. An overview of these techniques and their applications can be found in reviews [18, 19].

Information theory is of particular interest when it comes to deciphering the neural code [20]. Biological neural networks use statistical features of the

environment to efficiently encode sensory stimuli in the brain [21]. This encoding process can be studied in two ways.

The first one is by iteratively constructing optimal stimuli for specific neurons. This process, known as activation maximization, is used in artificial neural networks to interpret their inner workings [22, 8]. Existing approaches to creating optimal stimuli for biological neurons have led to unexpected and significant insights into neural coding. Examples include the complex structure of optimal stimuli for V1 neurons in the mouse primary visual cortex [23] and the so-called facial neurons [3], as well as the representation of relevant environmental elements in neuronal superstimuli [24]. Despite the promising results, this research area requires further development. The main problems are going beyond visual modality and the further development and scaling of computational algorithms to obtain optimal stimuli more quickly and efficiently [25, 26].

The second approach used in this study involves simultaneous recording of neural activity and various other time-series data, such as animal behavior, environmental conditions, and stimuli, among others. This data is then analyzed statistically to identify any stable relationships between the neural activity and these other variables. Although it has less flexibility in determining neural selectivity, since environmental and behavioral variables are calculated in advance, this approach allows for the analysis of the selectivity of a large number of neurons.

Prefrontal Neurons Encode a Solution to the Credit-Assignment Problem [27]

### 1.3 Calcium imaging and analysis of behavior-related neuronal activity

Calcium imaging

Importance of free behavior

Селективные нейроны могут формироваться быстро [28] [search](#)

There exist multiple methods for analyzing the relationship between neuronal spikes and behavioral variables ?. Some of them are based on an information-theoretical approach [29]. However, it is of particular interest to extract information about the relationship of the raw calcium fluorescent signal with behavior, since this signal provides important additional information about the subthreshold activations (?) of the neuron.

### 1.4 Methods of MI estimation

Mutual information can be used as a model-free measure of the relationship between time series, capable of quantifying non-linear interactions of the variables under study [30, 31].

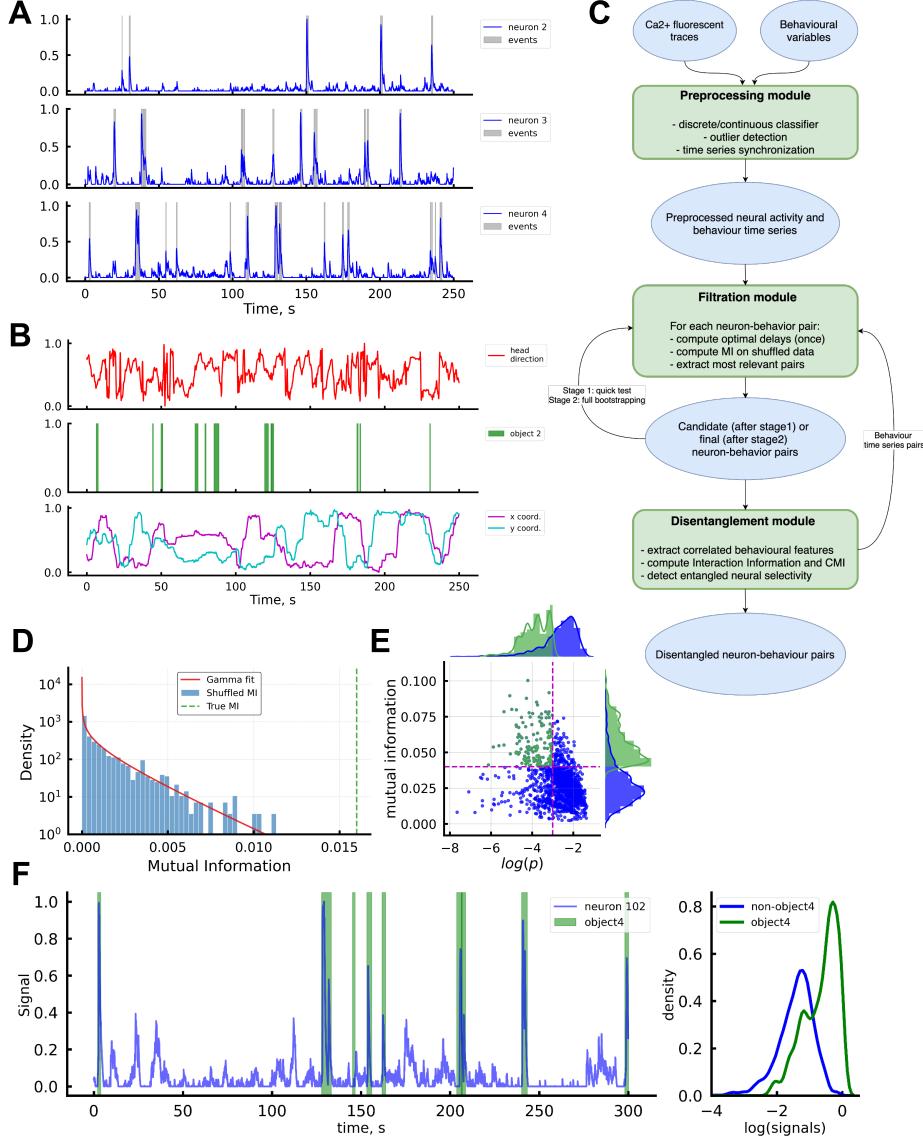


Рис. 1: **A:** Examples of rescaled  $dF/F$  calcium fluorescence traces. Grey regions represent detected calcium events. **B:** Examples of various behavioral/environmental variables to correlate neural activity with: continuous (head direction, top), discrete (interaction with object, middle), multidimensional (animal coordinates, bottom). **C:** Principal schema of INTENSE pipeline for neuronal selectivity quantification. **D:** Distribution of MI values for multiple shuffles (blue) along with the gamma distribution fit (red). True MI value for non-shuffled data is shown in green. **E:** distribution of neuron-behavior pairs in the “significance-power” coordinates and corresponding marginal distributions. Pairs considered relevant are shown in green. **F:** Left: an example of a detected “object interaction” cell: rescaled calcium signal is shown in blue, object interaction periods are shown in green. Right: distributions of scaled  $dF/F$  values inside (green) and outside (blue) the object interaction periods.

At the same time, such an analysis is complicated by the fact that the calcium signal, unlike discrete spikes, is a continuous quantity. Behavioral variables and environmental parameters can also be continuous (e.g. animal speed, pupil area, etc.) or discrete (rear, grooming) variables. The standard way to work with continuous variables in the information-theoretical approach is to divide the range of possible values into a pre-selected number of intervals. Thus, working with continuous variables is reduced to discrete ones, for which all information-theoretical quantities can be easily calculated. However, this approach suffers from serious drawbacks related to the arbitrariness of the number of intervals chosen [32]. To overcome these disadvantages, nonparametric methods have been developed [33, 32, 34], which are also used in this work.

Another difficulty lies in the unavoidable limited recording duration, which does not allow for a sufficiently complete sampling of the entire space of possible values. This means that in the conditions of a real neurobiological experiment, the estimation of information-theoretical quantities may contain a significant error. Because of this, the absolute values of MI expressed in bits may be inaccurate and do not reflect the relevant information content. In this regard, a “statistical” approach is more promising for the analysis of neural signals, in which the value of MI is considered not by itself, but as a random variable with some distribution. In this paradigm, the question “how much information does  $X$  contain about  $Y$ ” is replaced by “how significant is the information connection between  $X$  and  $Y$ ” ([35]).

while CI and MEA are both capable of measuring topology and dynamics of a neuronal network simultaneously, MEA consistently underestimates mutual information while CI may either underestimate or overestimate it (add spurious information) [36]

KSG estimator [33]

Mass univariate analysis of event-related brain potentials/fields [35]

Finally, the third difficulty lies in the large amount of calculations that are necessary both for the mass information analysis of pairs “fluorescent signal of neuron X - behavioral variable Y” and for constructing the default distribution of MI when calculating the significance of information communication. The large volume of neural signals recorded *in vivo*, combined with the need for statistical analysis, dictate strict performance requirements for the algorithms used.

Mutual information is copula entropy [37]

Про применения GCMI

In this paper, we propose a computational framework **INTENSE** (INformation-Theoretic Evaluation of Neuronal SElectivity) which effectively deals with the aforementioned problems.

Our contributions are as follows:

- We apply an information-theoretic approach to unravel the connections between neuronal calcium fluorescence signals and animal behavior variables, which allows us to characterize the nonlinear relationship between them. Our calculations are based on the fast and efficient Gaussian copula mutual information estimator (GCMI) [34], which we apply for the first time to calcium imaging data.
- We demonstrate the advantages of the information-theoretic approach (high robustness, small number of parameters, the ability to work in presence of high noise) and also propose a software tool capable of automatically searching for non-trivial relationships between neuronal activity and behavior: INTENSE.
- We show that INTENSE outperforms other commonly used methods on synthetic data, mimicking the structure of a real behavioral experiment. We also show its ability to recall correct neuron-behavior relationships even if the neuronal response is irregular and stochastic, which is essential for calcium imaging data.
- We verify our approach using *in vivo* data on the task of enriched environment exploration with a decreasing level of novelty in mice. We demonstrate the existence of hippocampal neurons which are selective for various aspects of animal behavior, including multiple neurons with mixed selectivity. We also propose a way to separate “true” mixed selectivity from secondary behavioral correlation effects (“selectivity disentanglement”).

## 2 Results

### 2.1 Comparison with classic methods of place cells detection

To verify the precision of INTENSE, we applied it to the widespread task of searching for place cells (PC), neurons that are selective for the location of the animal in the environment [38].

It is known that the choice of the place cell classification method greatly determines the population of cells identified [39], with overlap as low as 30-40% even for different event-based methods.

We compared the results of INTENSE, which was applied directly to calcium fluorescence signals and animal coordinates (see Methods: INTENSE PC identification) and the “classic” method based on the construction of a binarized activity map and the calculation of spatial information from computed calcium events [40] (see Methods: Classic PC identification)

Every PC computed by both methods was assigned an associated confidence measure ( $p$  value, see Methods: comparison of PC populations”). The results obtained show that, despite the overlap between two PC populations being modest

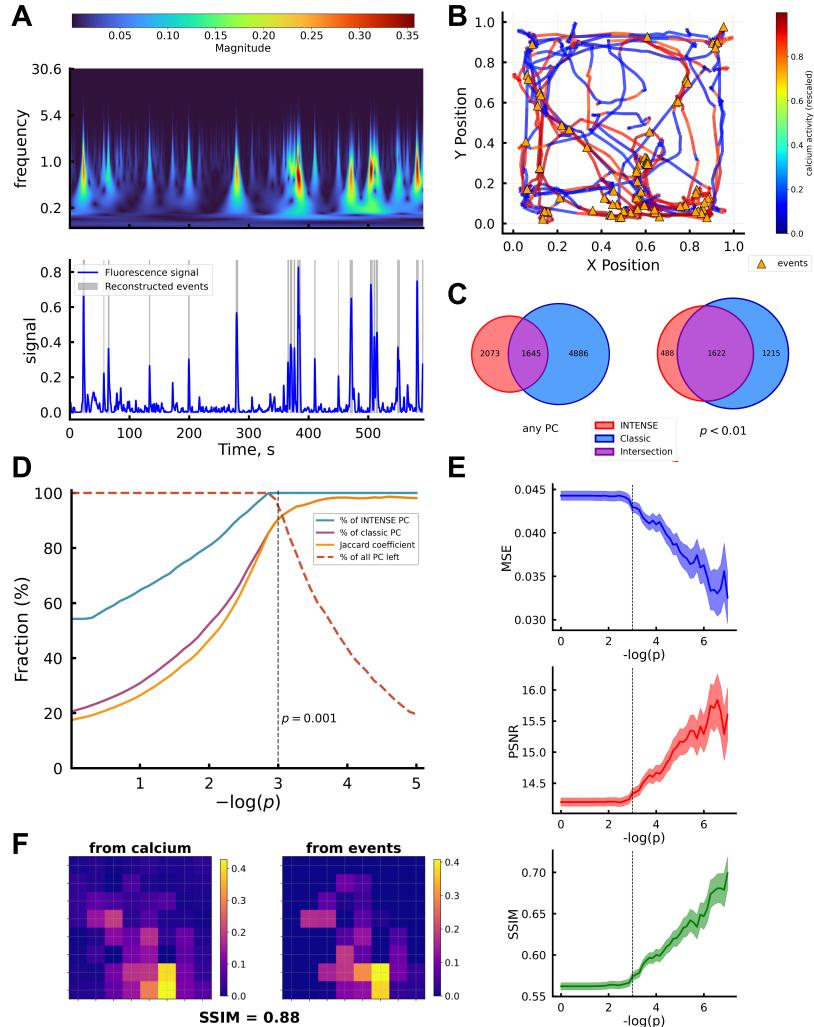


Рис. 2: **A:** Upper - CWT map of a fluorescence signal. Cross-scale ridges correspond to detected events. Lower - the signal itself (blue) and detected events (grey). **B:** An example animal trajectory, colors represent rescaled calcium activity of a single place-selective cell. Triangles represent detected events. **C:** Left - intersection of identified PC populations from INTENSE (red) and classic analysis (blue). All cells identified as PC in any method are shown. Right: the same for cells identified as PC with  $p < 0.01$  by both methods. **D:** Relative intersection between PC populations computed via INTENSE and classic analysis as a function of confidence (measured via  $p$ -value). Intersection is computed on cells for which  $p_{\text{INTENSE}} \leq p$  and  $p_{\text{classic}} \leq p$  simultaneously. Neurons considered as “non-PC” by both analyses are excluded. Blue: Intersection population as a fraction of INTENSE population. Shown are intersections as fractions of respective populations (blue for INTENSE, magenta for classic); Jaccard coefficient  $J = (\text{INTENSE} \cap \text{Classic}) / (\text{INTENSE} \cup \text{Classic})$ , orange; proportion of all identified PCs (by any method) left after thresholding by  $p$ . **E:** Similarity metrics between activity maps computed from raw calcium signals and from extracted events as functions of confidence (measured via  $p$ -value). Top: mean squared error, middle: peak signal-to-noise ratio, bottom: structural similarity score. Vertical lines show the  $p = 0.001$  threshold, as in Fig.2D. **F:** Example of activity maps computed from raw calcium signals (left) and from extracted events (right) with high SSIM score.

if all cells identified as PC by any method are taken, the overlap significantly increases if we constrain the  $p$ -value for both methods (see Fig. 2C). The overlap monotonically increases with the threshold, reaching 90% at a reasonable confidence threshold of  $p = 0.001$  (see Fig. 2D). At this point, only 5% of the total PC population is filtered out. A further increase of the threshold leads to almost 100% coincidence of both methods, at the cost of a substantial reduction of the PC population.

Significant heterogeneity of PC firing rates and stability has been established both experimentally [41] and theoretically [42]. Our results indicate that INTENSE and the spike-based method differ in how they classify “weak” place neurons, but converge on the same “core” PC population when subject to confidence constraints. Therefore, the general information-theoretic approach adapted by INTENSE successfully reproduces the results of a place-specific event-based method.

In order to additionally verify that all matching PCs identified by both methods have the same pattern of spatial selectivity, we analyzed the coincidence of activity maps. There exists a wide variety of place field detection methods [43], but they are highly dependent on the data type. To overcome this difficulty, we compared discretized activity maps (normalized by time spent in the bins; see Methods: Comparison of spatial selectivity and Fig. 2 F).

To address the similarity of activity maps, we computed several spatial correspondence metrics, including well-established (mean squared error, MSE and peak signal-to-noise ratio, PSNR) and a new one (structural similarity score, SSIM). Given the wide variety of image quality and comparison metrics [44], SSIM is a good proxy of a human eye similarity evaluation and is robust to noise. The example of such activity maps and their SSIM is presented in Fig. 2 F. The results show that all spatial correspondence metrics are growing the with confidence boundary (see 2 E), confirming that “strong” place cells indeed have more similar place fields obtained by both methods.

## 2.2 INTENSE deciphers neurons with mixed selectivity

Then we applied INTENSE to the task of finding links between recorded behavioral variables (see Methods: behavioral analysis) and neural activity. Additionally, after the standard procedure of searching for significant neuron-behavior pairs, we performed “disentangling” for neurons selective for similar variables using INTENSE behavior-behavior analysis and interaction information (see Fig. 3 D,E and Methods: Disentangling mixed neuronal selectivity).

The results showed the presence of multiple neurons that are selective for various specific behavioral acts (such as interacting with an object, rearing, etc., as shown in Fig. 3 A). These neurons are not always activated when the corresponding act occurs, and they can also be activated in the absence of the specific act. This can be explained by the probabilistic nature of neural coding

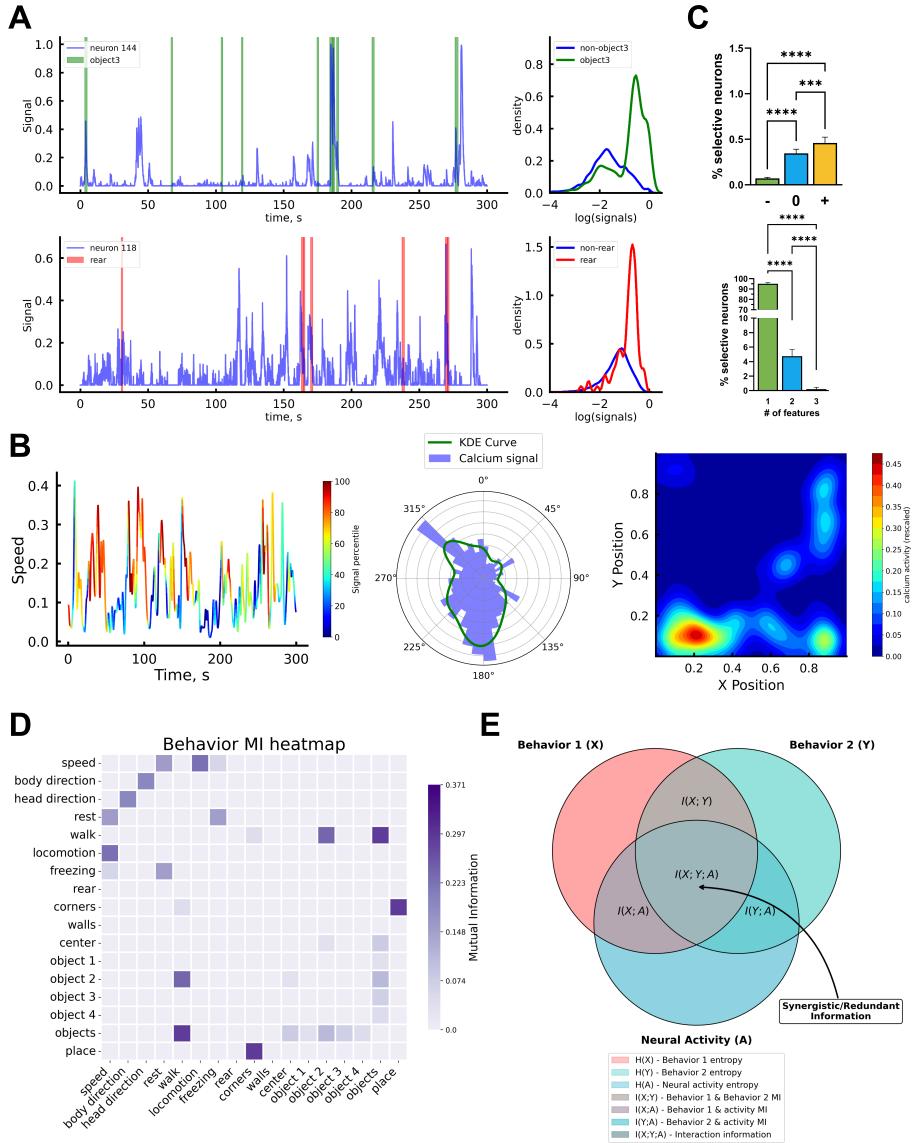


Рис. 3: **A:** Examples of neurons selective to discrete behavioral features found with INTENSE. Upper: “object interaction neuron”, lower: “rear neuron”. Graphs on the right show signal distributions inside respective behavior periods and outside them. **B:** Example of a multi-selective neuron: speed  $\times$  head direction  $\times$  place. Left: animal speed colored with calcium signal percentile values. Center: head direction tuning map of the same cell. Right: 2D activity map of this neuron. **C:** Upper - fraction of all recorded neurons, selective to some features with a negative (-), close to 0 and positive (+) delays. Lower - fraction of all selective neurons, which are selective to 1, 2 or 3 different features respectively. Statistical analysis was performed in GraphPad Prism 10.4.0. (GraphPad Software, USA) using one-factor ANOVA analysis of variance and Tukey’s a posteriori criterion. Differences were considered significant at  $p < 0.05$ . Data are presented as an average  $\pm$  95% confidence interval. **D:** Heatmap of mutual information between significantly correlated behavioral features, (computed via INTENSE behavior-behavior analysis). Darker color represents more significant connection. **E:** Schematic of information quantities relationship when dealing with “entangled” neuronal selectivity.

and the need for more detailed identification of behavioral acts based on the behavioral context. Acts that may seem identical from the perspective of the researcher based on the kinematic characteristics of the reference points, may have different internal meanings for the animal.

Along with the neurons of discrete behavioral acts, numerous cells have been found that are selective for continuous variables: speed, head direction, and, most often, position in space. [link to literature](#).

It is important to note that even after the disentanglement procedure, which eliminated the “parasitic” multiple selectivity, many neurons remained associated with more than one behavioral variable. For example, the entire neuron in Fig. 3 B is selective for three variables at once: the speed of the animal, the direction of the head, and the position in space. Despite the fact that these neurons are of great interest as points of intersection of different functional systems in the brain, their number is small compared to the total number of neurons (see Fig. 3 C).

For further statistical analysis of optimal delays between neural activity and behavior time series, the delay interval was divided into three equal parts, corresponding to negative, near-zero, and positive delays. For the analysis, only selective neurons (to some aspect of behavior or the environment) were selected.

For most of the behavioral variables analyzed, significant differences in the number of selective neurons with different time delays were shown (see Fig. 3 C). Thus, we concluded that selective neurons with negative delays make an insignificant contribution to the total pool of selective neurons. In this regard, we decided to consider only selective neurons isolated without or with positive time delays in further analysis.

### 2.3 INTENSE shows high performance on synthetic data

To better test the ability of INTENSE to reveal significant neuron-behavior relationships, we performed a series of tests on a synthetic dataset with *a priori* known “ground truth”.

We tested the ability to decipher neuron-behavior relationships for continuous and discrete synthetic behavioral variables separately (continuous and discrete test, respectively). For each test, we created 20 pseudo-behavioral time series  $b_i$  and 500 pseudo-calcium signals  $n_j$ , each of which was linked to exactly one of the  $b_i$  (see Fig.x and Methods: synthetic data generation). So, out of 10000 neuron-behavior pairs possible, 500 pairs were interlinked.

For the continuous test, we compared the following methods that exploit relationships between  $b_i$  and  $n_j$ :

- correlation-based
- MI-based

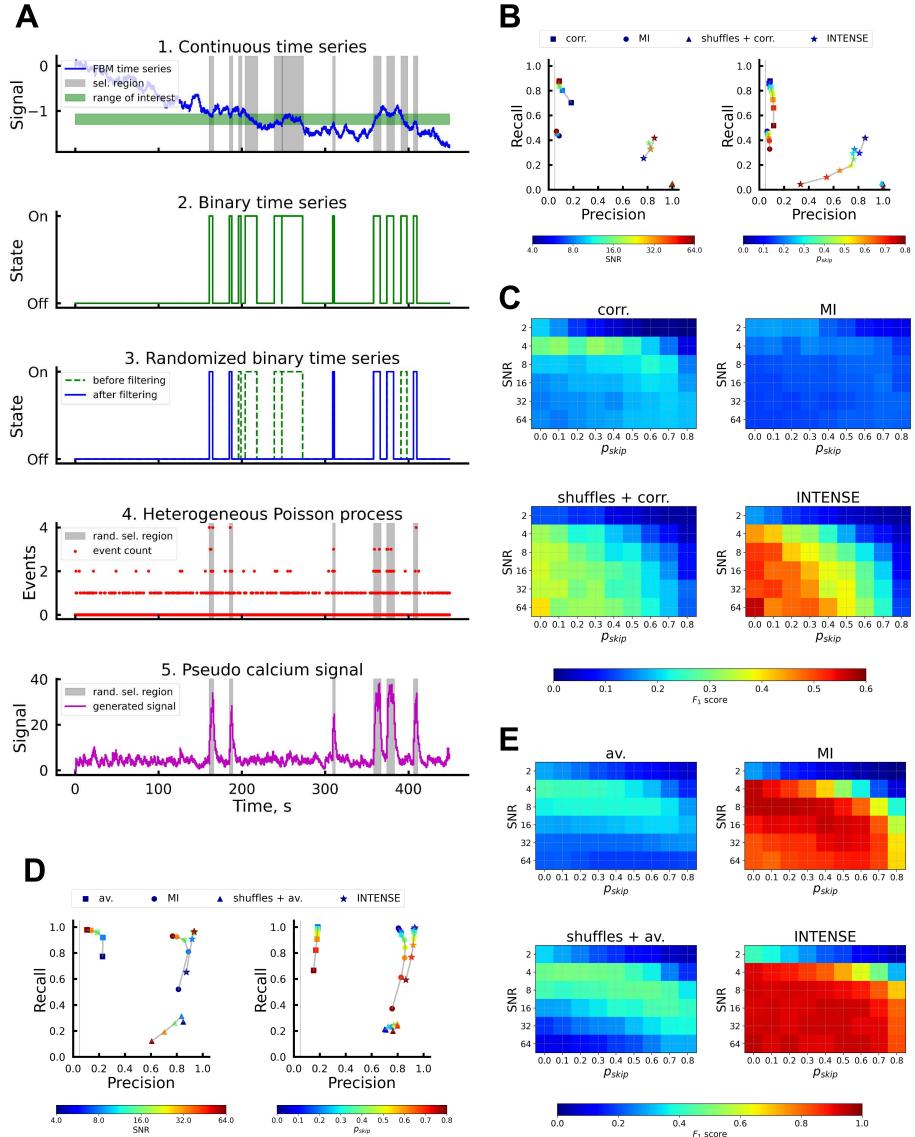


Рис. 4: **A:** Stages of generating pseudo-calculus signal, associated with a continuous variable. 1 - range of interest selection; 2 - creating binary time series; 3 - randomizing binary time series with  $p_{skip}$ ; 4 - Heterogeneous Poisson process with two different rates; 5 - Convolution with calcium signal-specific kernel. For discrete variables, the procedure starts from stage 2. **B, D:** Precision-Recall maps for continuous and discrete variable test, respectively. Methods shown for continuous test (B): correlation-based, MI-based, correlation-based with shuffles, INTENSE. Methods shown for discrete test (D): average-based, MI-based, average-based with shuffles, INTENSE. Left - precision-recall coordinates of all methods for different SNR values (taken at  $p_{skip} = 0$ ). Right - precision-recall coordinates of all methods for different  $p_{skip}$  values (taken at  $SNR = 64$ ). Vertical lines represent random guesser precision ( $PR_{random} = 0.05$ ). **C, E:**  $F_1$ -score heatmaps for the whole parameter range. C: continuous test, shown are the same methods as in B. E: discrete test, shown are the same methods as in D.

- correlation-based with shuffles
- INTENSE

For the discrete test, the methods used were as follows:

- average-based
- MI-based
- average-based with shuffles
- INTENSE

Such choice of methods provided the opportunity of an 'ablation study' of INTENSE. It allowed for quantitative analysis of contributions of nonlinear association measure (MI) and bootstrap procedure.

[results on  \$F\_1\$  and precision-recall](#)

## 3 Methods

### 3.1 Experimental setup

To search for selective neurons in mice, we chose a model of free behavior. The animals were repeatedly placed in a familiar environment. As a result, the level of novelty in this environment gradually decreased.

#### 3.1.1 Animals

TODO-KT, TODO-OI: [check this](#) Mice of the C57BL/6 line (both sexes), aged 3-5 months, were used in this study. The non-operated animals were housed 2-7 per cage, in laboratory cages, with free access to water and food, on a 12-hour light/dark cycle. After surgery, mice were housed individually, but other housing conditions did not change. All experiments were conducted during the light phase of the day, between 10:00 a.m. and 6:00 p.m., to minimize any potential circadian effects.

#### 3.1.2 Surgical procedures

[details here](#)

### 3.1.3 Behavioral analysis

Behavioral variables were extracted using the Sphynx package for exploratory activity analysis [45]. Initial body part positions were obtained frame-by-frame using a DeepLabCut-trained neural network [46]. During preprocessing, coordinates with confidence scores below 0.95 were excluded and reconstructed via piecewise cubic interpolation based on adjacent reliable points. Time series were then smoothed using a third-degree Savitzky-Golay filter, with a sliding window of 0.25 seconds for the animal's body center and tail base, and 0.1 seconds for other body parts.

For each video recording, the experimental arena was divided into distinct spatial zones: a wall zone (7 cm from the walls), a central zone (30×30 cm square), corner zones (7×7 cm from the walls), and object zones (2.5 cm perimeter around objects). Behavioral variables—both continuous and discrete—were derived by analyzing the relative positions of the animal's body parts with respect to these zones.

Continuous variables included Cartesian coordinates (defined by the body center position), absolute velocity (computed from the body center speed and further smoothed with a 0.25-second Savitzky-Golay filter), body direction (BD, the angle of the vector from the body center to the head center), and head direction (HD, the angle of the vector from the head center to the nose tip).

Discrete behavioral acts were classified based on movement dynamics and spatial positioning. Locomotion bouts were classified by absolute velocity: fast locomotions - when animal's speed exceeded 5 cm/s, slow locomotions - when animal's speed was between 1-5 cm/s, and rests periods - when speed did not exceed 1 cm/s. Freezing was identified when both body center velocity remained below 1 cm/s and nose tip velocity did not exceed 2 cm/s. Rearing was detected by a reduction in the distance between the hind limbs and tail base. Space-related acts were registered when the body center entered predefined zones (corners, walls, or center), while object-related acts were triggered by the nose tip entering object perimeters. All discrete acts were subjected to median smoothing over a 0.25-second window, corresponding to the minimal plausible duration of a behavioral act.

### 3.1.4 $Ca^{2+}$ imaging

TODO-VP TODO-OR

[check this](#)

The activity of neurons in the CA1 region of the hippocampus was recorded using a miniature microscope (miniscope, Miniscope V4.4). To do this, the animals underwent stereotactic surgery, in which a viral vector carrying the GCaMP6s fluorescent calcium sensor gene was introduced, a GRIN lens (1 mm) was implanted and miniscope mounts were installed. Two weeks after the

operation, the mice were trained to connect and wear a miniscope for 5 minutes for three days in a home cage. Then, for 4 days, the mice were placed in an open field for 10 minutes (OP, Fig. 9), in which 4 types of objects are located in the center of the arena, and visual landmarks are placed on the walls. The calcium activity of the CA1 region was recorded on all days of the examination by OP mice. Animal behavior was recorded using a video camera (Flir Chameleon3), synchronization of two data streams was carried out in the Bonsai software environment.

### 3.1.5 Fluorescent image analysis

TODO-VP: [check this section](#)

The analysis of calcium activity was performed using the software package BEARMIND developed by the authors (see Code Availability). This software is based on the CaImAn package [47] and is specifically designed for end-to-end multi-session analysis of miniscope calcium activity videos. A primary visual analysis of the video recordings was performed and the spatial cropping parameters were determined. The preprocessing of video data included... ([DeepCad?](#)). Next, the correction of motion artifacts was performed based on the NoRMCorre algorithm [48]. The processed video signal was then decomposed into a spatial mask, which shows the probable neuron-related ROIs, and the calcium activity time series of these ROIs using the constrained nonnegative matrix factorization (CNMF-E) technique [49].

The algorithm hyperparameters were selected by a human expert by inspecting the corresponding processed images. The major parameters influencing the CNMF decomposition results were selected as follows:

- gSig, reflecting the characteristic size of the neuron, was selected in the range of 3-6 pixels (resolution 3 microns/pixel)
- min\_corr, reflecting the density of the population of the field of view by neurons, was selected in the range 0.85-0.95
- min\_pnr determining the lower bound in the signal-to-noise ratio for neuron activity was selected from 3 to 15.
- After signal factorization, the components were automatically selected according to the threshold rval\_thr=0.95 and min\_SNR=3 (according to CaImAn guidelines).

All selected components were subjected to human expert inspection to identify artifacts and duplicate components, which was performed with the help of BEARMIND software. Further, all components (and their temporal and spatial decompositions) that were selected and inspected were considered to correspond to neurons. The time series of the calcium response was normalized in the form  $dF/F$ , and the spatial components were combined between sessions using the CellReg procedure [50].

### 3.1.6 Event detection

Calcium events were extracted from  $dF/F$  calcium fluorescence signal following the algorithm described in [51] with minor improvements. The wavelet-based algorithm was selected because of its noise robustness and superior performance on potentially noisy data. First, the signal was smoothed by convolution with a Gaussian kernel with  $\sigma = 8$ . Then, a continuous wavelet transform (CWT) was applied to the signal, resulting in a two-dimensional table of wavelet coefficients containing information about “ridges” of local maxima at all spatial scales (see Fig. ??A). Significant events can be identified by the properties of these ridges. We used a generalized Morse wavelet  $\Psi_{\beta,\gamma}(\omega) = \alpha_{\beta,\gamma}\omega^\beta e^{-\omega^\gamma}$  with shape parameter  $\gamma = 3$  and initial scale parameter  $\beta = 2$ . Wavelets with  $\gamma = 3$ , also called Airy wavelets, were chosen because they have the smallest Heisenberg area.

All local maxima of CWT amplitude were determined for each scale. These local maxima serve to initialize new ridges and continue existing ones. Local maxima of CWT amplitude at the largest scale size initialize ridges. When transitioning to smaller scales, local maxima were merged into a ridge if they fall within a temporal range defined by the central window for the wavelet of the previous scale centered on the ridge. If several smaller-scale maxima fell within the range defined for the ridge, only the largest of the new maxima was added to the ridge. Other elements in the window initialized new ridges. If a local maximum did not fall into any of the windows of previous ranges or maxima, it also initialized a new ridge. The algorithm terminated either after processing maxima at the smallest scale, when all local maxima at all scales are assembled into ridges (or identified as isolated), or after reaching a user-specified scale. Gaps in ridges were not allowed - if a maximum does not exist at some scale, the ridge was terminated. Once all ridges in the CWT decomposition were identified, they underwent significance filtering (thresholds were applied for maximum duration, amplitude, as well as the scale at which the maximum value in the ridge was observed). The filtered ridges represented calcium events.

## 3.2 INTENSE pipeline

### 3.2.1 MI computation

Neural selectivity analysis in INTENSE is conducted by calculating mutual information between neural signal and animal behavior. Mutual information was chosen as the metric due to its ability to account for nonlinear interactions between the studied variables. INTENSE employs the Gaussian Copula Mutual Information (GCMI) method for computing mutual information [34]. The method is based on the fact that mutual information between two random variables is independent of their marginal distributions, but depends only on the type of copula (a multivariate distribution where each marginal distribution is uniform).

Here we briefly recall the details of the GCMI method.

For two random variables  $X$  and  $Y$ , the mutual information is defined as:

$$I(X; Y) = H(X) + H(Y) - H(X, Y) \quad (1)$$

The GCMI method transforms the data to a Gaussian copula representation:

$$\tilde{X} = \Phi^{-1}(F_X(X)), \quad \tilde{Y} = \Phi^{-1}(F_Y(Y)) \quad (2)$$

where  $F_X$  and  $F_Y$  are the empirical cumulative distribution functions, and  $\Phi^{-1}$  is the inverse standard normal CDF.

### Continuous-Continuous Case

For two continuous variables, GCMI estimates mutual information as:

$$I_{GCMI}(X; Y) = -\frac{1}{2} \log |\mathbf{R}| \quad (3)$$

where  $\mathbf{R}$  is the correlation matrix of the copula-transformed variables  $(\tilde{X}, \tilde{Y})$ .

### Discrete-Continuous Case

For a discrete variable  $X$  with  $K$  states and continuous variable  $Y$ :

$$I_{GCMI}(X; Y) = \sum_{k=1}^K p(X = k) \cdot I(X = k; Y|X = k) \quad (4)$$

where each conditional term is computed using the continuous GCMI formula on the subset of data where  $X = k$ .

For each significant selectivity, its strength was determined as the mutual information between the fluorescent signal and behavior:

$$S = I(X_{neural}; Y_{behavior}) \quad (5)$$

#### 3.2.2 Significance Testing

To assess the significance of the calculated informational connection between calcium signal and behavioral variable, we compared the true MI with values calculated on randomly time-shifted signals (shuffles):

$$MI_{shuffle}^{(i)} = I(X(t); Y(t + \tau_i)) \quad (6)$$

where  $\tau_i$  represents random circular shifts.

The null distribution of shuffled MI values is modeled using a pre-selected parametric distribution, with gamma distribution as the default choice:

$$MI_{shuffle} \sim \Gamma(\alpha, \beta) \quad (7)$$

where parameters  $\alpha$  (shape) and  $\beta$  (rate) are estimated from the empirical shuffle distribution using maximum likelihood estimation.

The significance threshold is then established by comparing  $MI_{true}$  against this fitted distribution:

$$p\text{-value} = P(MI_{shuffle} \geq MI_{true}) = 1 - F_\Gamma(MI_{true}; \hat{\alpha}, \hat{\beta}) \quad (8)$$

where  $F_\Gamma$  is the cumulative distribution function of the fitted gamma distribution.

This parametric approach provides a smoother estimate of extreme tail probabilities compared to the empirical distribution, particularly important when the number of shuffles is limited.

$$\log \frac{p(x, y)}{p(x)p(y)} \quad (9)$$

### 3.2.3 Computing optimal delays

The calcium fluorescence signal has a characteristic dynamics that is slightly delayed from the immediate spike or a bundle of spikes of a neuron. The magnitude of this delay is determined by several factors, among which the main ones are the parameters of the calcium sensor used, and is usually 1-2 s [52].

However, even in the absence of delay behavior and neural activity are not necessarily rigidly correlated with each other over time. For example, it is known that there are neurons encoding the distance from a goal, whose activity correlates with its future achievement [53]. The activity of some neurons may be triggered by the onset of a behavioral act, which will lead to its observed delay relative to the behavioral variable. To account for these effects, we add the ability to calculate the significance of the information link between neural activity and time-delayed behavior.

To do this, a delay is selected in the range of  $[-2, 2]$  seconds in increments of 0.05 seconds, maximizing mutual information between the selected neural signal and behavior. After that, the same calculations are performed as in the case without taking into account delays, with the difference that the “true” mutual information between the signals is considered to be information calculated with an optimal delay between them.

### 3.2.4 Selecting significant neuron-behavior relations

The choice of MI distribution type is an important consideration. Empirical distributions of MI have a heavy tail and therefore cannot be accurately approximated by a normal distribution. Moreover, since MI is inherently non-negative, its distribution should take into account this property. Gamma and log-normal

distributions are two possible options that satisfy these requirements and are supported by theoretical research [54].

We performed selectivity analysis using a two-stage procedure that allowed us to increase the accuracy and statistical significance of results.

1. **First stage:** We calculated 100 random shuffles for each neuron-behavioral variable pair, comparing the mutual information (MI) values in shuffles with the true value. A pair was considered to have passed the first phase if the true MI was greater than in all shuffles:

$$MI_{true} > \max_{i=1}^{100} MI_{shuffle}^{(i)} \quad (10)$$

This criterion is necessary for early rejection of unpromising pairs, allowing us to substantially reduce computational load on one hand, and decrease the number of potential hypotheses at the multiple comparisons correction stage on the other.

2. **Second stage:** We conducted a more precise comparison with 10,000 shuffles, applying two criteria:

- (a) The true MI exceeds the 99.95th percentile of the random MI distribution:

$$MI_{true} > Q_{99.95}(MI_{shuffle}) \quad (11)$$

- (b) The  $p$ -value of a one-sided  $t$ -test for the hypothesis of equality between the mean shuffle distribution and true MI does not exceed a given threshold:

$$p = P(t > t_{obs}) < p_{threshold} \quad (12)$$

$$\text{where } t_{obs} = \frac{MI_{true} - \overline{MI}_{shuffle}}{SE(MI_{shuffle})}$$

The threshold  $p$ -value was determined using the Holm correction for multiple comparisons with the number of hypotheses equal to the number of pairs that passed the first stage, and a family-wise error rate (FWER) of 0.01:

$$p_{threshold}^{(k)} = \frac{\alpha}{m - k + 1} \quad (13)$$

where  $\alpha = 0.01$ ,  $m$  is the total number of hypotheses, and  $k$  is the rank of the ordered  $p$ -values.

After conducting the significance analysis, we obtained neuron-variable pairs for which the mutual information between the calcium signal and behavior was significantly higher than chance. However, many such pairs were characterized by very weak synchronization effects between calcium activity and behavior that could not be verified by expert examination. These phantom” specializations could be artifacts of short experimental recording length or caused by single

“external” synchronizing events, such as sound or illumination changes. Despite formally demonstrated statistical significance of this synchronization, it is useless for further analysis. Therefore, we applied an additional criterion for connection strength:

$$MI_{true} > MI_{threshold} \quad (14)$$

where  $MI_{threshold}$  is an absolute value common for all sessions of a given experiment.

### 3.2.5 Disentangling mixed neuronal selectivity

Some pairs of behavioral variables added to the automatic analysis inevitably turn out to be significantly correlated with each other. For example, the binary variables “locomotion” and “stops” are obviously interrelated, as well as the continuous variables “head direction” and “body direction”. Consequently, the neural selectivities for these variables identified using INTENSE will also be interrelated and will often appear together in further analysis. To assess how much such pairwise selectivities are artifacts of the chosen variables, as a first step, we applied INTENSE directly to pairs of behavioral variables (see Fig. ?? D). This analysis was conducted analogously to the “neuron-behavior” pair analysis.

If there was no informational connection between two behavioral variables, we considered that all neurons selective for both of them possess “true” multi-selectivity, i.e., participate in the operation of two different functional systems.

In case of identifying a significant connection between two behavioral variables, we analyzed neurons selective for both these variables using more complex information-theoretic tools that allow determining the nature of their relationships.

For a more precise assessment of interactions in the triangle of variables “neural activity - behavior 1 - behavior 2” we calculated the conditional mutual information between these variables (see Fig. ?? E).

In what follows, we adopt the notation  $A$  - neural activity variable,  $X$  - behavioral variable 1,  $Y$  - behavioral variable 2.

There are 4 possible scenarios in total:

1.  $X, Y$  - continuous

1. In this case, the conditional mutual information  $I(A, X|Y)$  is calculated using the formula:

$$I(A, X|Y) = H(A, Y) + H(X, Y) - H(A, X, Y) - H(Y)$$

Individual entropy components are estimated using the GCMI method, through Cholesky decomposition of covariance matrices of data transformed to normal form. 2.  $X$  - continuous,  $Y$  - discrete

2. In this case, the conditional mutual information  $I(A, X|Y)$  is calculated separately for each discrete value of variable  $Y$  using the GCMI method through the formula:

$$I(A, X) = H(A) + H(X) - H(A, X)$$

The obtained values are then weighted with the probabilities of the corresponding values of  $Y$ . 3.  $Y$  - continuous,  $X$  - discrete

3. In this case, the conditional mutual information  $I(A, X|Y)$  is calculated from the equality:

$$I(A, X|Y) = I(A, X) - (I(A, Y) - I(A, Y|X))$$

The terms  $I(A, X)$  and  $I(A, Y)$  are calculated using the GCMI method for a combination of discrete and continuous variables. The term  $I(A, Y|X)$  is calculated as in point 2. 4.  $X, Y$  - discrete

4. In this case, the conditional mutual information  $I(A, X|Y)$  is calculated as the sum:

$$I(A, X|Y) = H(A, Y) + H(X, Y) - H(A, X, Y) - H(Y)$$

The term  $H(A, Y)$  is calculated analogously to point 1 separately for all values of  $Y$ , with subsequent weighting by the probabilities of occurrence of various values of  $Y$ . The terms  $H(X, Y)$  and  $H(Y)$  are calculated using the standard entropy formula for two-dimensional or one-dimensional discrete quantities, respectively. The term  $H(A, X, Y)$  is calculated through the sum of conditional entropies for each possible pair of values  $X, Y$ .

Using the conditional mutual information, we calculated the so-called interaction information (II) - a quantity characterizing the synergetic/redundant relationship between three random variables. II characterizes the influence of the third variable  $Y$  on information transfer between  $A$  and  $X$  and is calculated using the formula:

$$I(A, X, Y) = I(A, X) - I(A, X|Y)$$

The value of II does not depend on the order of variables and in practice is calculated by averaging over several different orders.

In further analysis, we use the result of [55], where the authors considered the interaction of three variables in terms of a directed acyclic graph and connected it with the II value. In particular, in the case of negative II (the most frequently encountered case in practice when working with neural data), it is known that only one of the connections between the three variables can be "weak" (the numerical criterion for "weakness" of an informational connection is that its numerical value is less than the modulus of II). In this case, only two interaction graphs between the variables are possible, which are shown in Fig. 5.

Thus, the further analysis is organized as follows: 1. The II value between the three variables is calculated 2. The strengths of informational connections

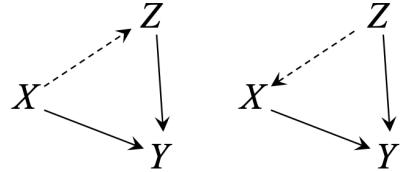


Рис. 5: caption (taken from [55]: two possible acyclic interaction graphs between the three considered variables under the condition of weak connection between A and Y.

between pairs  $AX$ ,  $AY$  (neural activity with both behavioral variables) are compared with II value 3. In case of a weak connection, the other two are automatically considered strong - in this case, the behavioral variables turn out to be "nested" one within the other, and we can consider that only the strong connection is meaningful (case 1). 4. In case there is no weak connection, it is considered that the behavioral variables have equal strength, in which case they are most likely substantially duplicated, and the decision about which one to keep should be made by an expert (case 2).

Examples of the first situation are the variables speed (continuous) and locomotion (discrete). The analysis shows that neurons selective for both variables possess selectivity for the presence of locomotion itself, while speed becomes a "secondary" factor.

An example of the second situation can be the relationship between the variables "head direction" and "body direction". Since there is a high correlation between them, information-theoretic methods cannot resolve the question of which is more important, and we must recognize them as equivalent. In practice, one of them should be discarded as duplicative.

### 3.3 Place cells analysis

#### 3.3.1 Classic PC identification

Place cell identification was performed using a previously described custom MATLAB routine [56]. The arena space was divided into square bins measuring  $8 \times 8$  cm. For each neuron exhibiting at least  $n = 5$  calcium events, an activity map was generated by dividing the smoothed number of calcium events by the smoothed occupancy time for each bin. The resulting activity map was segmented into distinct regions using a watershed thresholding approach, and the spatial information content for each region was calculated. To assess the significance of spatial tuning, 1000 surrogate datasets were created by randomly shifting calcium activity traces relative to the animal's trajectory, and the spatial information content was recomputed for each shuffled trial. The p-value for the

observed spatial information was determined based on this distribution. Regions where spatial information exceeded the 95th percentile of the shuffled distribution were classified as candidate place fields. Within these informative regions, place fields were defined as subareas where the neuron firing rate exceeded 50% of the maximum firing rate in the corresponding informative area.

check and extend TODO-VP: is this part ok?

### 3.3.2 INTENSE PC identification

To quantify neuronal coordinate selectivity with INTENSE, we have performed the analysis described in TODO, using the joint animal coordinates  $C = X, Y$  as a behavioral variable. Therefore, all information-theoretic quantities were computed for joint coordinates, including entropy  $H(C) = H(X, Y)$  and mutual information  $I(A, C) = I(A; X, Y)$ . When shuffling the data, components of  $C$  were shifted together by the same delay.

### 3.3.3 Comparison of PC populations

### 3.3.4 Comparison of spatial selectivity

## 3.4 Synthetic data

### 3.4.1 Feature generation

### 3.4.2 Signal generation

## 4 Discussion

spikes: A comparison of neuronal population dynamics measured with calcium imaging and electrophysiology [57]

Quantifying uncertainty in spikes estimated from calcium imaging data [58]

INTENSE is specifically aimed to uncover neurons related to aspects of the external environment or the animal's behavior. However, it can also be used as a tool for unraveling nonlinear relationships between multiple time series, for example, constructing functional neuronal connectome.

multivariate IT

При этом большинство нейронов отзываются на нелинейный сложный микс task-related variables [59, 60]. Судя по всему, это является неизбежным следствием высокой размерности "нейронного резервуара". Такой взгляд на специализации также хорошо согласуется с log-dynamic brain гипотезой

[61], согласно которой небольшая часть самых сильно специализированных нейронов стабильна, а специализации остальных постоянно плавают и трансформируются опытом. Например, сила и устойчивость специализаций положительно коррелируют с firing rates нейронов в гиппокампе [62]. Пластичность специализаций большинства клеток может быть основой для nonlinear mixed selectivity, а она позволяет устойчиво кодировать информацию о многих переменных в мозге [63]. Глубокие сети также демонстрируют mixed selectivity при кодировании, причем чем она больше, тем лучше performance [64].

better way of representing causality: Estimating transfer entropy via copula entropy [65]

Brain-like functional specialization emerges spontaneously in deep neural networks [66] (на примере face)

Мультиомадальные нейроны концептов в мозге [67] и трансформерах [68]. Эпизодические нейроны в гиппокампе людей [69]. Современные LLM тоже формируют специализированные нейроны, например “space neurons” and “time neurons” [? ]

Возбуждение единственной клетки из ансамбля способно вызвать массовые эффекты и влиять на поведение [70]

Изменение поведения отдельных селективных нейронов также может влиять на работу больших языковых моделей [71].

## 5 Data availability

Raw miniscope videos, fluorescence traces, detected events, and computation results are available upon reasonable request.

## 6 Code availability

Our analysis was performed using open-source software developed at the Institute for Advanced Brain Studies, MSU.

The BEARMIND pipeline for miniscope video analysis is available at <https://github.com/iabs-neuro/bearmind>

The SPHYNX (Segmented PHYSical aNalysis of eXploration) pipeline for animal behavior analysis is available at <https://github.com/iabs-neuro/sphynx>

TODO-VP: внести полные координаты Сфинкса и Бирмайнда сюда

[DRIADA](#)

[GCMI code](#) [34]

Python package ssqueezepy [72] designed for wavelet analysis and synchrosqueezing was used to compute fast and accurate signal CWT with generalized Morse wavelets.

## 7 Acknowledgements

This work was supported by Non-Commercial Foundation for Support of Science and Education "INTELLECT" and [+funding](#). We are grateful to Sergey Nечаev and Olga Martynova for valuable comments and discussions.

## Список литературы

- [1] D. H. Hubel, T. N. Wiesel, Brain and visual perception: the story of a 25-year collaboration, Oxford University Press, 2004.
- [2] S. P. O Scalaidhe, Face-selective neurons during passive viewing and working memory performance of rhesus monkeys: Evidence for intrinsic specialization of neuronal coding, *Cerebral Cortex* 9 (5) (1999) 459–475. doi:10.1093/cercor/9.5.459.  
URL <http://dx.doi.org/10.1093/cercor/9.5.459>
- [3] A. Bardon, W. Xiao, C. R. Ponce, M. S. Livingstone, G. Kreiman, Face neurons encode nonsemantic features, *Proceedings of the National Academy of Sciences* 119 (16) (Apr. 2022). doi:10.1073/pnas.2118705119.  
URL <http://dx.doi.org/10.1073/pnas.2118705119>
- [4] R. Quijan Quiroga, M. Boscaglia, J. Jonas, H. G. Rey, X. Yan, L. Maillard, S. Colnat-Coulbois, L. Koessler, B. Rossion, Single neuron responses underlying face recognition in the human midfusiform face-selective cortex, *Nature Communications* 14 (1) (Sep. 2023). doi:10.1038/s41467-023-41323-5.  
URL <http://dx.doi.org/10.1038/s41467-023-41323-5>
- [5] P. Montes-Lourido, M. Kar, S. V. David, S. Sadagopan, Neuronal selectivity to complex vocalization features emerges in the superficial layers of primary auditory cortex, *PLOS Biology* 19 (6) (2021) e3001299. doi:10.1371/journal.pbio.3001299.  
URL <http://dx.doi.org/10.1371/journal.pbio.3001299>
- [6] B. Roland, T. Deneux, K. M. Franks, B. Bathellier, A. Fleischmann, Odor identity coding by distributed ensembles of neurons in the mouse olfactory cortex, *eLife* 6 (May 2017). doi:10.7554/eLife.26337.  
URL <http://dx.doi.org/10.7554/eLife.26337>

- [7] C. M. A. Pennartz, M. N. Oude Lohuis, U. Olcese, How ‘visual’ is the visual cortex? the interactions between the visual cortex and other sensory, motivational and motor systems as enabling factors for visual perception, *Philosophical Transactions of the Royal Society B: Biological Sciences* 378 (1886) (Aug. 2023). doi:10.1098/rstb.2022.0336.  
 URL <http://dx.doi.org/10.1098/rstb.2022.0336>
- [8] G. Goh, N. Cammarata, C. Voss, S. Carter, M. Petrov, L. Schubert, A. Radford, C. Olah, Multimodal neurons in artificial neural networks, *Distill* 6 (3) (2021) e30.
- [9] W. J. Johnston, S. Fusi, Abstract representations emerge naturally in neural networks trained to perform multiple tasks, *Nature Communications* 14 (1) (Feb. 2023). doi:10.1038/s41467-023-36583-0.  
 URL <http://dx.doi.org/10.1038/s41467-023-36583-0>
- [10] C. E. Shannon, A mathematical theory of communication, *Bell System Technical Journal* 27 (3) (1948) 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.  
 URL <http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- [11] A. I. Luppi, E. A. Stamatakis, Combining network topology and information theory to construct representative brain networks, *Network Neuroscience* 5 (1) (2021) 96–124. doi:10.1162/netn\_a\_00170.  
 URL [http://dx.doi.org/10.1162/netn\\_a\\_00170](http://dx.doi.org/10.1162/netn_a_00170)
- [12] T. F. Varley, M. Pope, J. Faskowitz, O. Sporns, Multivariate information theory uncovers synergistic subsystems of the human cerebral cortex, *Communications Biology* 6 (1) (Apr. 2023). doi:10.1038/s42003-023-04843-w.  
 URL <http://dx.doi.org/10.1038/s42003-023-04843-w>
- [13] S. Nigam, M. Shimono, S. Ito, F.-C. Yeh, N. Timme, M. Myroshnychenko, C. C. Lapish, Z. Tosi, P. Hottowy, W. C. Smith, S. C. Masmanidis, A. M. Litke, O. Sporns, J. M. Beggs, Rich-club organization in effective connectivity among cortical neurons, *The Journal of Neuroscience* 36 (3) (2016) 670–684. doi:10.1523/jneurosci.2177-15.2016.  
 URL <http://dx.doi.org/10.1523/JNEUROSCI.2177-15.2016>
- [14] N. Kannathal, M. L. Choo, U. R. Acharya, P. Sadasivan, Entropies for detection of epilepsy in eeg, *Computer Methods and Programs in Biomedicine* 80 (3) (2005) 187–194. doi:10.1016/j.cmpb.2005.06.012.  
 URL <http://dx.doi.org/10.1016/j.cmpb.2005.06.012>
- [15] F. von Wegner, H. Laufs, Information-theoretical analysis of eeg microstate sequences in python, *Frontiers in Neuroinformatics* 12 (Jun. 2018). doi:10.3389/fninf.2018.00030.  
 URL <http://dx.doi.org/10.3389/fninf.2018.00030>

- [16] J. G. Orlandi, O. Stetter, J. Soriano, T. Geisel, D. Battaglia, Transfer entropy reconstruction and labeling of neuronal connections from simulated calcium imaging, *PLoS ONE* 9 (6) (2014) e98842. doi:10.1371/journal.pone.0098842.  
 URL <http://dx.doi.org/10.1371/journal.pone.0098842>
- [17] A. S. Blevins, D. S. Bassett, E. K. Scott, G. C. Vanwalleghem, From calcium imaging to graph topology, *Network Neuroscience* 6 (4) (2022) 1125–1147. doi:10.1162/netn\_a\_00262.  
 URL [http://dx.doi.org/10.1162/netn\\_a\\_00262](http://dx.doi.org/10.1162/netn_a_00262)
- [18] A. G. Dimitrov, A. A. Lazar, J. D. Victor, Information theory in neuroscience, *Journal of Computational Neuroscience* 30 (1) (2011) 1–5. doi:10.1007/s10827-011-0314-3.  
 URL <http://dx.doi.org/10.1007/s10827-011-0314-3>
- [19] N. M. Timme, C. Lapish, A tutorial for information theory in neuroscience, *eneuro* 5 (3) (2018) ENEURO.0052–18.2018. doi:10.1523/eneuro.0052-18.2018.  
 URL <http://dx.doi.org/10.1523/ENEURO.0052-18.2018>
- [20] A. Borst, F. E. Theunissen, Information theory and neural coding, *Nature Neuroscience* 2 (11) (1999) 947–957. doi:10.1038/14731.  
 URL <http://dx.doi.org/10.1038/14731>
- [21] H. B. Barlow, et al., Possible principles underlying the transformation of sensory messages, *Sensory communication* 1 (01) (1961) 217–233.
- [22] D. Bau, B. Zhou, A. Khosla, A. Oliva, A. Torralba, Network dissection: Quantifying interpretability of deep visual representations, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 6541–6549.
- [23] E. Y. Walker, F. H. Sinz, E. Cobos, T. Muhammad, E. Froudarakis, P. G. Fahey, A. S. Ecker, J. Reimer, X. Pitkow, A. S. Tolias, Inception loops discover what excites neurons most using deep predictive models, *Nature neuroscience* 22 (12) (2019) 2060–2065.
- [24] C. R. Ponce, W. Xiao, P. F. Schade, T. S. Hartmann, G. Kreiman, M. S. Livingstone, Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences, *Cell* 177 (4) (2019) 999–1009.
- [25] W. Xiao, G. Kreiman, Gradient-free activation maximization for identifying effective stimuli, *arXiv preprint arXiv:1905.00378* (2019).
- [26] N. Pospelov, A. Chertkov, M. Beketov, I. Oseledets, K. Anokhin, Fast gradient-free activation maximization for neurons in spiking neural networks (2024). doi:10.48550/ARXIV.2401.10748.  
 URL <https://arxiv.org/abs/2401.10748>

- [27] W. F. Asaad, P. M. Lauro, J. A. Perge, E. N. Eskandar, Prefrontal neurons encode a solution to the credit-assignment problem, *The Journal of Neuroscience* 37 (29) (2017) 6995–7007. doi:10.1523/jneurosci.3311-16.2017.  
 URL <http://dx.doi.org/10.1523/JNEUROSCI.3311-16.2017>
- [28] V. P. Sotskov, N. A. Pospelov, V. V. Plusnin, K. V. Anokhin, Calcium imaging reveals fast tuning dynamics of hippocampal place cells and ca1 population activity during free exploration task in mice, *International Journal of Molecular Sciences* 23 (2) (2022) 638. doi:10.3390/ijms23020638.  
 URL <http://dx.doi.org/10.3390/ijms23020638>
- [29] A. Borst, F. E. Theunissen, Information theory and neural coding, *Nature neuroscience* 2 (11) (1999) 947–957.
- [30] A. Dionisio, R. Menezes, D. A. Mendes, Mutual information: a measure of dependency for nonlinear time series, *Physica A: Statistical Mechanics and its Applications* 344 (1–2) (2004) 326–329. doi:10.1016/j.physa.2004.06.144.  
 URL <http://dx.doi.org/10.1016/j.physa.2004.06.144>
- [31] X. Zhao, P. Shang, J. Huang, Mutual-information matrix analysis for nonlinear interactions of multivariate time series, *Nonlinear Dynamics* 88 (1) (2016) 477–487. doi:10.1007/s11071-016-3254-7.  
 URL <http://dx.doi.org/10.1007/s11071-016-3254-7>
- [32] B. C. Ross, Mutual information between discrete and continuous data sets, *PLoS ONE* 9 (2) (2014) e87357. doi:10.1371/journal.pone.0087357.  
 URL <http://dx.doi.org/10.1371/journal.pone.0087357>
- [33] A. Kraskov, H. Stögbauer, P. Grassberger, Estimating mutual information, *Physical Review E* 69 (6) (Jun. 2004). doi:10.1103/physreve.69.066138.  
 URL <http://dx.doi.org/10.1103/PhysRevE.69.066138>
- [34] R. A. Ince, B. L. Giordano, C. Kayser, G. A. Rousselet, J. Gross, P. G. Schyns, A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula, *Human Brain Mapping* 38 (3) (2016) 1541–1573. doi:10.1002/hbm.23471.  
 URL <http://dx.doi.org/10.1002/hbm.23471>
- [35] D. M. Groppe, T. P. Urbach, M. Kutas, Mass univariate analysis of event-related brain potentials/fields i: A critical tutorial review, *Psychophysiology* 48 (12) (2011) 1711–1725. doi:10.1111/j.1469-8986.2011.01273.x.  
 URL <http://dx.doi.org/10.1111/j.1469-8986.2011.01273.x>
- [36] T. Jung, F. Vogiatzian, O. Har-Shemesh, C. Fitzsimons, R. Quax, Applying information theory to neuronal networks: From theory to experiments, *Entropy* 16 (11) (2014) 5721–5737. doi:10.3390/e16115721.  
 URL <http://dx.doi.org/10.3390/e16115721>

- [37] J. Ma, Z. Sun, Mutual information is copula entropy, *Tsinghua Science & Technology* 16 (1) (2011) 51–54.
- [38] J. O’Keefe, J. Dostrovsky, The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat., *Brain research* (1971).
- [39] D. M. Grijseels, K. Shaw, C. Barry, C. N. Hall, Choice of method of place cell classification determines the population of cells identified, *PLOS Computational Biology* 17 (7) (2021) e1008835. doi:10.1371/journal.pcbi.1008835.  
URL <http://dx.doi.org/10.1371/journal.pcbi.1008835>
- [40] W. Skaggs, B. Mcnaughton, K. Gothard, An information-theoretic approach to deciphering the hippocampal code, *Advances in neural information processing systems* 5 (1992).
- [41] K. Mizuseki, S. Royer, K. Diba, G. Buzsáki, Activity dynamics and behavioral correlates of ca3 and ca1 hippocampal pyramidal neurons, *Hippocampus* 22 (8) (2012) 1659–1680. doi:10.1002/hipo.22002.  
URL <http://dx.doi.org/10.1002/hipo.22002>
- [42] Y. Lian, A. N. Burkitt, Learning spatiotemporal properties of hippocampal place cells, *eneuro* 9 (4) (2022) ENEURO.0519–21.2022. doi:10.1523/eneuro.0519-21.2022.  
URL <http://dx.doi.org/10.1523/ENEURO.0519-21.2022>
- [43] R. M. Grieves, K. J. Jeffery, The representation of space in the brain, *Behavioural Processes* 135 (2017) 113–131. doi:10.1016/j.beproc.2016.12.012.  
URL <http://dx.doi.org/10.1016/j.beproc.2016.12.012>
- [44] U. Sara, M. Akter, M. S. Uddin, Image quality assessment through fsim, ssim, mse and psnr—a comparative study, *Journal of Computer and Communications* 07 (03) (2019) 8–18. doi:10.4236/jcc.2019.73002.  
URL <http://dx.doi.org/10.4236/jcc.2019.73002>
- [45] V. Plusnin, O. Ivashkina, N. Pospelov, O. Rogozhnikova, N. Savelev, V. Sotskov, K. Toropova, K. Anokhin, Sphynx: An automated behavioral analysis tool for neuronal selectivity identification, in: 2024 Sixth International Conference Neurotechnologies and Neurointerfaces (CNN), IEEE, 2024, p. 156–159. doi:10.1109/cnn63506.2024.10705840.  
URL <http://dx.doi.org/10.1109/CNN63506.2024.10705840>
- [46] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, M. Bethge, Deeplabcut: markerless pose estimation of user-defined body parts with deep learning, *Nature Neuroscience* 21 (9) (2018) 1281–1289. doi:10.1038/s41593-018-0209-y.  
URL <http://dx.doi.org/10.1038/s41593-018-0209-y>

- [47] E. A. Pnevmatikakis, D. Soudry, Y. Gao, T. A. Machado, J. Merel, D. Pfau, T. Reardon, Y. Mu, C. Lacefield, W. Yang, M. Ahrens, R. Bruno, T. M. Jessell, D. S. Peterka, R. Yuste, L. Paninski, Simultaneous denoising, deconvolution, and demixing of calcium imaging data, *Neuron* 89 (2) (2016) 285–299. doi:10.1016/j.neuron.2015.11.037.  
 URL <http://dx.doi.org/10.1016/j.neuron.2015.11.037>
- [48] E. A. Pnevmatikakis, A. Giovannucci, Normcorre: An online algorithm for piecewise rigid motion correction of calcium imaging data, *Journal of Neuroscience Methods* 291 (2017) 83–94. doi:10.1016/j.jneumeth.2017.07.031.  
 URL <http://dx.doi.org/10.1016/j.jneumeth.2017.07.031>
- [49] P. Zhou, S. L. Resendez, J. Rodriguez-Romaguera, J. C. Jimenez, S. Q. Neufeld, A. Giovannucci, J. Friedrich, E. A. Pnevmatikakis, G. D. Stuber, R. Hen, M. A. Kheirbek, B. L. Sabatini, R. E. Kass, L. Paninski, Efficient and accurate extraction of *in vivo* calcium signals from microendoscopic video data, *eLife* 7 (Feb. 2018). doi:10.7554/elife.28728.  
 URL <http://dx.doi.org/10.7554/elife.28728>
- [50] L. Sheintuch, A. Rubin, N. Brande-Eilat, N. Geva, N. Sadeh, O. Pinchasof, Y. Ziv, Tracking the same neurons across multiple days in  $\text{Ca}^{2+}$  imaging data, *Cell Reports* 21 (4) (2017) 1102–1115. doi:10.1016/j.celrep.2017.10.013.  
 URL <http://dx.doi.org/10.1016/j.celrep.2017.10.013>
- [51] A. Neugornet, B. O'Donovan, P. I. Ortinski, Comparative effects of event detection methods on the analysis and interpretation of  $\text{Ca}^{2+}$  imaging data, *Frontiers in Neuroscience* 15 (Mar. 2021). doi:10.3389/fnins.2021.620869.  
 URL <http://dx.doi.org/10.3389/fnins.2021.620869>
- [52] Y. Zhang, M. Rózsa, Y. Liang, D. Bushey, Z. Wei, J. Zheng, D. Reep, G. J. Broussard, A. Tsang, G. Tsegaye, S. Narayan, C. J. Obara, J.-X. Lim, R. Patel, R. Zhang, M. B. Ahrens, G. C. Turner, S. S.-H. Wang, W. L. Korff, E. R. Schreiter, K. Svoboda, J. P. Hasseman, I. Kolb, L. L. Looger, Fast and sensitive gcamp calcium indicators for imaging neural populations, *Nature* 615 (7954) (2023) 884–891. doi:10.1038/s41586-023-05828-9.  
 URL <http://dx.doi.org/10.1038/s41586-023-05828-9>
- [53] M. J. Beetz, C. Kraus, B. el Jundi, Neural representation of goal direction in the monarch butterfly brain, *Nature Communications* 14 (1) (Sep. 2023). doi:10.1038/s41467-023-41526-w.  
 URL <http://dx.doi.org/10.1038/s41467-023-41526-w>
- [54] M. Hutter, Distribution of mutual information, *Advances in neural information processing systems* 14 (2001).

- [55] A. Ghassami, N. Kiyavash, Interaction information for causal inference: The case of directed triangle (2017). [arXiv:1701.08868](https://arxiv.org/abs/1701.08868).  
 URL <https://arxiv.org/abs/1701.08868>
- [56] V. Plusnin, O. Ivashkina, K. Toropova, V. Sotskov, A. Tiaglik, K. Anokhin, Neuronal encoding of objects and place in hippocampus: the value of objects shapes memory, in: 2021 Third International Conference Neurotechnologies and Neurointerfaces (CNN), IEEE, 2021, p. 83–86. doi:[10.1109/cnn53494.2021.9580384](https://doi.org/10.1109/CNN53494.2021.9580384).  
 URL <http://dx.doi.org/10.1109/CNN53494.2021.9580384>
- [57] Z. Wei, B.-J. Lin, T.-W. Chen, K. Daie, K. Svoboda, S. Druckmann, A comparison of neuronal population dynamics measured with calcium imaging and electrophysiology, *PLOS Computational Biology* 16 (9) (2020) e1008198. doi:[10.1371/journal.pcbi.1008198](https://doi.org/10.1371/journal.pcbi.1008198).  
 URL <http://dx.doi.org/10.1371/journal.pcbi.1008198>
- [58] Y. T. Chen, S. W. Jewell, D. M. Witten, Quantifying uncertainty in spikes estimated from calcium imaging data, *Biostatistics* 24 (2) (2021) 481–501. doi:[10.1093/biostatistics/kxab034](https://doi.org/10.1093/biostatistics/kxab034).  
 URL <http://dx.doi.org/10.1093/biostatistics/kxab034>
- [59] M. Rigotti, O. Barak, M. R. Warden, X.-J. Wang, N. D. Daw, E. K. Miller, S. Fusi, The importance of mixed selectivity in complex cognitive tasks, *Nature* 497 (7451) (2013) 585–590. doi:[10.1038/nature12160](https://doi.org/10.1038/nature12160).  
 URL <http://dx.doi.org/10.1038/nature12160>
- [60] S. Fusi, E. K. Miller, M. Rigotti, Why neurons mix: high dimensionality for higher cognition, *Current Opinion in Neurobiology* 37 (2016) 66–74. doi:[10.1016/j.conb.2016.01.010](https://doi.org/10.1016/j.conb.2016.01.010).  
 URL <http://dx.doi.org/10.1016/j.conb.2016.01.010>
- [61] G. Buzsáki, K. Mizuseki, The log-dynamic brain: how skewed distributions affect network operations, *Nature Reviews Neuroscience* 15 (4) (2014) 264–278. doi:[10.1038/nrn3687](https://doi.org/10.1038/nrn3687).  
 URL <http://dx.doi.org/10.1038/nrn3687>
- [62] K. Mizuseki, G. Buzsáki, Preconfigured, skewed distribution of firing rates in the hippocampus and entorhinal cortex, *Cell Reports* 4 (5) (2013) 1010–1021. doi:[10.1016/j.celrep.2013.07.039](https://doi.org/10.1016/j.celrep.2013.07.039).  
 URL <http://dx.doi.org/10.1016/j.celrep.2013.07.039>
- [63] W. J. Johnston, S. E. Palmer, D. J. Freedman, Nonlinear mixed selectivity supports reliable neural computation, *PLOS Computational Biology* 16 (2) (2020) e1007544. doi:[10.1371/journal.pcbi.1007544](https://doi.org/10.1371/journal.pcbi.1007544).  
 URL <http://dx.doi.org/10.1371/journal.pcbi.1007544>
- [64] Y. Wu, Mixed selectivity via unsupervised learning in neural networks, Ph.D. thesis (2019). doi:[10.17863/CAM.53941](https://doi.org/10.17863/CAM.53941).  
 URL <https://www.repository.cam.ac.uk/handle/1810/306850>

- [65] J. Ma, Estimating transfer entropy via copula entropy, arXiv preprint arXiv:1910.04375 (2019).
- [66] K. Dobs, J. Martinez, A. J. Kell, N. Kanwisher, Brain-like functional specialization emerges spontaneously in deep neural networks, *Science advances* 8 (11) (2022) eabl8913.
- [67] R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, I. Fried, Invariant visual representation by single neurons in the human brain, *Nature* 435 (7045) (2005) 1102–1107. doi:10.1038/nature03687.  
URL <http://dx.doi.org/10.1038/nature03687>
- [68] G. Goh, N. Cammarata, C. Voss, S. Carter, M. Petrov, L. Schubert, A. Radford, C. Olah, Multimodal neurons in artificial neural networks, *Distill* 6 (3) (Mar. 2021). doi:10.23915/distill.00030.  
URL <http://dx.doi.org/10.23915/distill.00030>
- [69] L. D. Kolibius, F. Roux, G. Parish, M. Ter Wal, M. Van Der Plas, R. Chelvarajah, V. Sawlani, D. T. Rollings, J. D. Lang, S. Gollwitzer, K. Walther, R. Hopfengärtner, G. Kreiselmeyer, H. Hamer, B. P. Staresina, M. Wimber, H. Bowman, S. Hanslmayr, Hippocampal neurons code individual episodic memories in humans, *Nature Human Behaviour* 7 (11) (2023) 1968–1979. doi:10.1038/s41562-023-01706-6.  
URL <http://dx.doi.org/10.1038/s41562-023-01706-6>
- [70] T. Alejandro-García, S. Kim, J. Pérez-Ortega, R. Yuste, Intrinsic excitability mechanisms of neuronal ensemble formation, *eLife* 11 (May 2022). doi:10.7554/elife.77470.  
URL <http://dx.doi.org/10.7554/elife.77470>
- [71] A. Templeton, T. Conerly, J. Marcus, J. Lindsey, T. Bricken, B. Chen, A. Pearce, C. Citro, E. Ameisen, A. Jones, H. Cunningham, N. L. Turner, C. McDougall, M. MacDiarmid, C. D. Freeman, T. R. Sumers, E. Rees, J. Batson, A. Jermyn, S. Carter, C. Olah, T. Henighan, Scaling monosemanticity: Extracting interpretable features from claude 3 sonnet, *Transformer Circuits Thread* (2024).  
URL <https://transformer-circuits.pub/2024/scaling-monosemanticity/index.html>
- [72] J. Muradeli, ssqueezepy, GitHub. Note: <https://github.com/OverLordGoldDragon/ssqueezepy/> (2020). doi:10.5281/zenodo.5080508.