2023

Two-way interactions in cortical computations. — In a collaborative work with the University of California, Los Angeles (UCLA) we have identified a novel approach for studying the neural code. Employing a machine learning methodology, we established that despite the distinctive "mixed selectivity" characteristic of individual neurons, it is possible to accurately read out a variety of environmental stimuli and even the cognitive state of animals. The study presenting these results was published in Nature Communications [1].
One fundamental insight of neuroscience is that nerve cells act as "experts" in supporting perception and decision-making. Certain cells are sensitive to various, even highly complex, characteristics of the environment, and through their activity, they can signal to later stages in the processing chain what the animal is currently facing.

Subsequent research, however, have revealed that an individual neuron is not as much of a specialist expert as previously thought, but nerve cells actually engage in enthusiastic "multitasking." This phenomenon, known as "mixed selectivity," may initially appear practical: the role of a single nerve cell is not limited to just one task, thus potentially increasing the capacity of neural circuits in the brain. However, mixed selectivity can pose a challenge for both the processing layers that read the output of nerve cells and analytical tools trying to decode the signals of nerve cells since it's not clear that the intensity of the nerve cell's response reflects the presence of specific environmental characteristics.

In our paper, however, we pointed out that despite mixed selectivity, various environmental stimuli, and even the cognitive state of animals can be clearly determined. Using a machine learning approach that served as the basis for the results, we determined that visual and auditory stimuli simultaneously encoded by nerve cells, as well as the context reflecting the regularities of a task, can be unambiguously identified as a simple geometric structure without interfering with each other. This solution can be imagined as the different faces of a cube, each storing separately the seen or the heard information as well as the current rule determining how the visual and auditory information shall be used (Figure 1). This cube can float in space in any direction, so we are not limited to one side, meaning we can perceive not only visual or auditory information from just one perspective. However, we can find a direction where we perceive only one type of information without interference from the other sides.

In a follow-up study we demonstrated how the identified representational geometry can be used to design optimal actions under challenging situations, in particular when simultaneously observed stimuli are in conflict with respect to the desired action [2]. we showed that neural activity in the Anterior Cingulate Cortex displays activity that is hallmark of attentional gating. Signatures of the identified Context-Dependent Gating mechanism was shown in model neural networks and in mice.

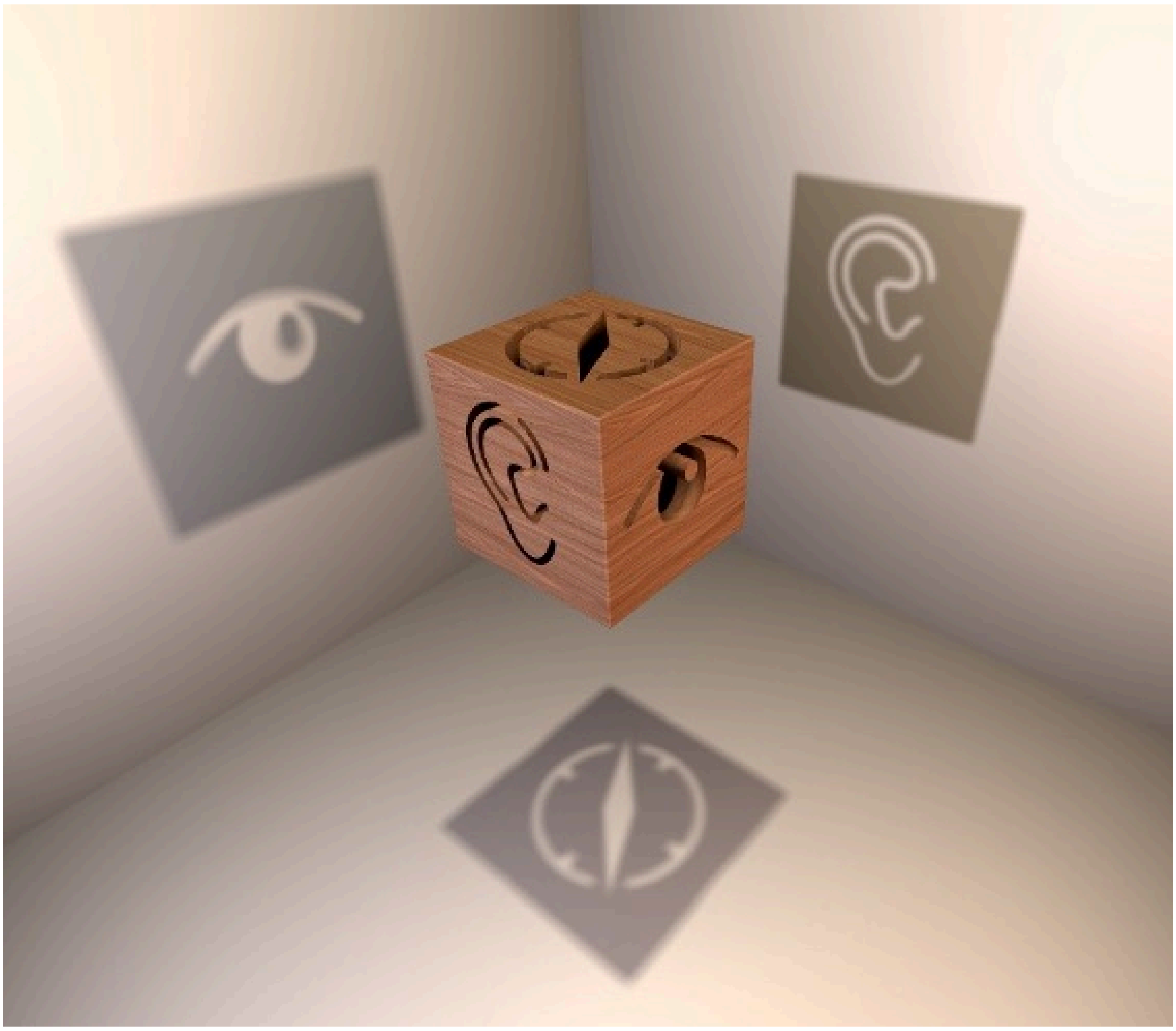

Figure 1. Illustration of the high-dimensional neural code that represents the wide spectrum of variables relevant for task execution. The variables are not necessarily represented on individual axes, which would correspond to a representational scheme where the variable is represented by a single neuron, but a viewing angle can be identified from which reading out the variable is simple.

References:
[1] Márton Albert Hajnal, Duy Tran, Karen Safaryan, Michael Einstein, Mauricio Vallejo Martelo, Pierre-Olivier Polack, Peyman Golshani, Gergő Orbán. Continuous multiplexed population representations of task context in the mouse primary visual cortex. Nat Commun 14, 6687 (2023). https://doi.org/10.1038/s41467-023-42441-w
[2] Márton Albert Hajnal, Duy Tran, Zsombor Szabó, Andrea Albert, Karen Safaryan, Michael Einstein, Mauricio Vallejo Martelo, Pierre-Olivier Polack, Peyman Golshani, Gergő Orbán. Shifts in attention drive context-dependent subspace encoding in anterior cingulate cortex during decision making. bioRxiv 2023.10.10.561737; doi: https://doi.org/10.1101/2023.10.10.561737

2021

Flexible control in the brain. — A key asset of biological intelligence is the ability to flexibly interpret environmental stimuli to design appropriate actions. A core challenge is that the same stimulus might have different consequences in different contexts and therefore invoke different actions. If the context is explicit, i.e. it is also cued by some sensory stimulus, then this problem is rather straightforward. If the context is not cued then the challenge becomes more interesting and it is unclear how the brain copes with it. One of the challenges is that while in the case of explicitly cued context only associations between sensory variables needs to be learned, in the uncued case context can only be inferred from trial history and needs to be maintained across trials. One way to do that is to maintain a routing from stimulus-reward contingencies how various stimuli needs to be mapped to decisions. Alternatively, the animal could learn to represent the unobserved context variable in the neural circuitry. Another challenge concerns the representation of stimulus. It is unknown whether input stimuli are represented in an invariant fashion and differently combined when the context changes or the stimulus representation itself undergoes transformations across contexts.

In a collaboration with Peyman Golshani's lab at the Medical School of University of California Los Angeles, we investigated mice using a paradigm in which auditotory and visual stimuli led to different rewards depending on the context. More precisely, mice were trained to execute a pair of tasks: in one reward depended on the content of visual stimulus, in the other it depended on the content of auditory stimulus.

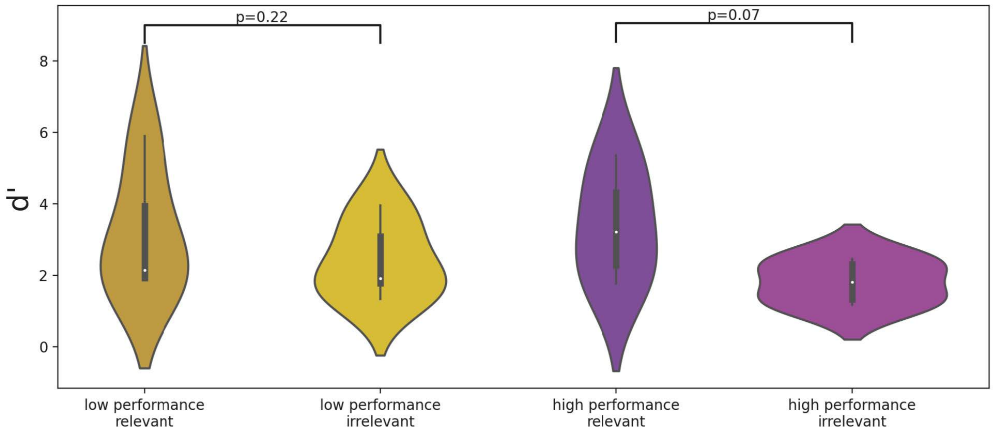

Figure 1. Specificity of neuron population responses to stimulus identity in the Anterior Cyngulate Cortex region of the mouse cortex. Relevant stimuli are represented with higher accuracy than irrelevant stimuli (purple distributions). However, this difference is contingent on whether the animal was well aware of the actual task being performed, as signalled by lower suppression of irrelevant stimuli when behavioral performance was deteriorating.

Previously, we have investigated the visual cortex, the area of the brain dedicated to handling visual information. We did not find signatures of context-dependent stimulus representation but could identify a context signal that was orthogonal to the visual stimulus and maintained across-trial information about the reigning context.

To pursue this line of study, we investigated if at later stages of study the representation of stimuli becomes affected by the context. We have shown that in an area implicated in higher level control, the Anterior Cyngulate Cortex, the representation becomes task-specific: when a stimulus is relevant, it is represented with higher fidelity while when it is irrelevant, the neural responses become less specific to that particular modality. we have also shown that this context relevant suppression of irrelevant signals correlates with behavior: in periods when the animal did not perform the task well, context-relevant suppression could not be identified in population responses. Our research provides fresh insights into the way animals flexibly infer the rules of the game even when the rules cannot be determined from the actual input but needs to be infered from task history.

References:

[1] Márton Albert Hajnal, Duy Tran, Michael Einstein, Mauricio Vallejo Martelo, Karen Safaryan, Pierre-Olivier Polack, Peyman Golshani, Gergő Orbán. Continuous multiplexed population representations of task context in the mouse primary visual cortex, biorxiv, 2021, doi:10.1101/2021.04.20.440666
[2] Hajnal MA, Szabó Z, Tran D, Einstein M, Martelo MV, Safaryan K, Polack P, Golshani P, Orbán G (2021) Anterior cingulate cortex represents cue relevance in a context-shifting task. Bernstein Conference 2021. doi: 10.12751/nncn.bc2021.p178

2020

Lossy compression in semantic memory. — Human memory is a surprising unreliable device. It is prone to let us down when we would like to remember our experiences and such failures can have dramatic consequences, such as when jurisdiction is relying on witness testimony. Yet, when discussing the brain as the result of a long optimization process, namely that of evolution, we tend to describe it as an optimal device. To assess if this is a contradiction, or the functioning of the memory is optimal in some respect we investigated the structure of memory errors. We use computer vision as an inspiration: storage of information in artificial information systems is also prone to commit errors, when capacity for storage is limited. The mathematical theory for compressing information is formulated in the rate distortion theory. We used the theory of lossy compression to understand and predict the memory errors that humans are making. Importantly we used inspiration from the differences between errors committed by computer vision technologies and those by the human memory. By using state-of-the-art machine learning to understand the statistical structure in naturalistic data we could show that a wide variety of human memory errors can be understood as optimally compressing data for efficient storage of information when resources are limited. Our research links a number of studies that found puzzling infidelities of human memory and has strong predictions on the way experiences are transformed as time passes. It also provides us with insights into the principles for efficient compression solutions.

References:

[1] Nagy D, Török B, Orbán G (2020) Optimal forgetting: Semantic compression of episodic memories. PLoS Computational Biology, 16(10): e1008367.

2019

How to read the neural code. — What are the principles underlying the computations performed by neuronal networks? Answering this question is central to neuroscience and underlies both the characterization of mental diseases and devices establishing interactions between humans and machines through neuronal signals. The work at the Computational systems neuroscience group investigated how single neuron responses and joint responses of pairs of neurons provide insights in this fundamental question.

Representational untangling in the visual cortex. — The visual cortex in the brain contributes to the interpretation of stimuli detected by our retinae. In order to interpret the environment subsequent stages of processing need a code that is easy to read, similar to the easily readable message we receive on our phone that we happily read as opposed to the gibberish sequence of zeros and ones that is communicated by our phone to the communication tower. The magic happening between these two codes is called nonlinear transformation. Such nonlinear transformations are certainly taken place in the brain but what these are and how these contribute to an easily readable code is a subject of debate.

In a collaboration with Peyman Golshani (UCLA), Máté Lengyel (University of Cambridge), and Pierre-Olivier Polack (Ruthgers University) we investigated the role of the most basic but most widespread form of nonlinearity, the so called firing rate nonlinearity, which determines how the level of membrane potential of the neuron is transformed into the intensity of action potential generation. We showed that a key characteristics of the nonlinearity, the threshold above which neurons start to generate action potentials, controls the readability of the neural code (Figure 1). We pointed out that the threshold controls two opposing effects. Lowered threshold results in increased information transfer, while elevated threshold contributes to increased sparseness of the code, meaning that responses in a population become more selective to specific stimuli, thus contributing to an easy to read dictionary. The result of the two opposing effects is that the best readable code corresponds to a sweet spot where both of these effects can exert their benefit. This is where the so-called representational untangling, the magic of nonlinear computation appears. Using recordings from neurons in mice we could demonstrate that the properties of recorded neurons is optimized for representational untangling, that is, an easy to read code.

Generative computations in the visual cortex. — Machine vision, and in particular so-called deep discriminative models, which are optimized to perform image categorization, have proven effective in describing how neurons respond to complex stimuli. Theoretical considerations, however, indicate that computations optimized to perform a single task do not reflect the flexibility of the nervous system, that is, its capability to build up representations serving for multiple tasks. Inspired by this insight, we designed an experiment that can demonstrate the presence of stimulus-dependent correlations, a hallmark of another class of machine vision models, generative models. As opposed to deep discriminative

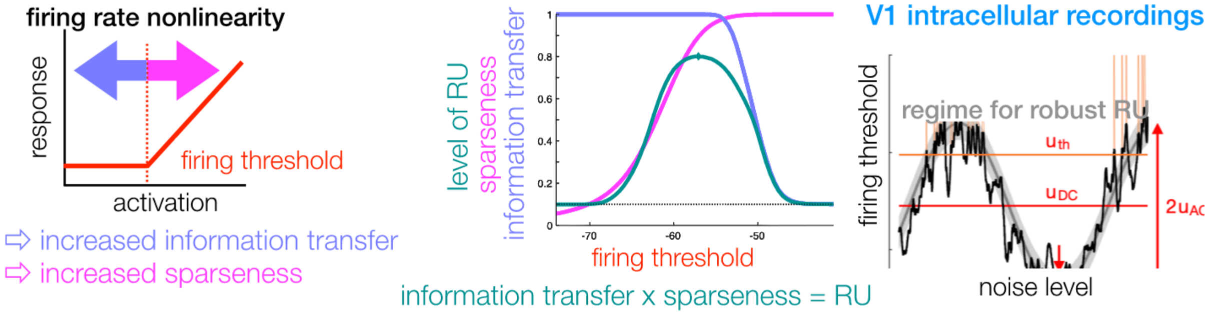

Figure 1. Left: Mapping of membrane potential activation to the intensity of action potential generation. Arrows denote potential changes in firing threshold. Middle: Changes in the fidelity of information transfer and sparseness as firing threshold is changed. Representational untangling (RU) is a result of the combined effect of information transfer and sparseness. Right: the parameter regime where RU is most effective (gray area) and measured properties of neurons in the visual cortex of mice. Overlap between the two indicates that neurons are optimized to achieve RU.

models, deep generative models build up a task-independent representation of the environment. Based on collaborations with Wolf Singer (MPI/ESI Frankfurt) we designed a set of behaving macaque experiments that could identify these hallmarks. Our results are based on the prediction that stimulus statistics affects the level of stimulus specificity of correlated variability between neurons (Figure 2). Critically, these results resolve a long standing puzzle in neuroscience. Traditional models, such as the models underlying deep discriminative models regard correlations as detrimental for efficient encoding of stimulus identity but these correlations are ubiquitous in the nervous system. Our results show that these correlations are a consequence of optimal computations.

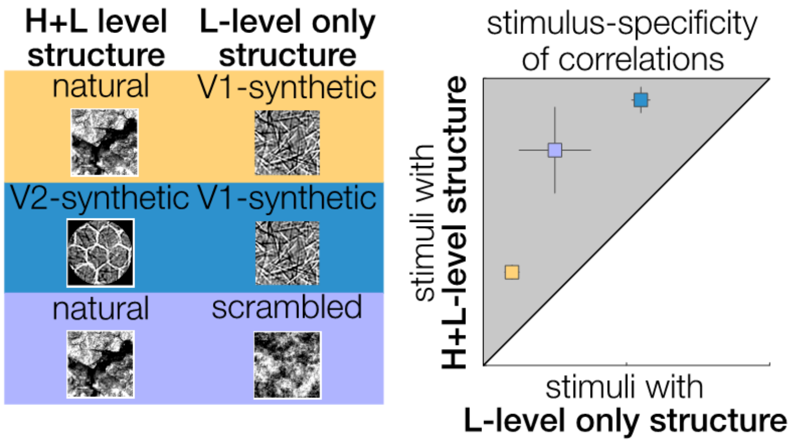

Figure 2. Stimulus-specificity of correlations. Deep generative models predict that images featuring high-level structure (left) display higher levels of stimulus-specificity in correlated variability than images only characterized by low-level structure (right).

External links:
[1] https://doi.org/10.1016/j.conb.2019.09.002
[2] https://doi.org/10.7554/eLife.43625
[3] https://doi.org/10.1073/pnas.1816766116

2017

Information processing in the visual system. In order to understand how the brain handles visual information, we build mathematical models of the visual cortex, and compare predictions derived from them to experimental data. Assessing how variability is introduced to neuronal firing is a prerequisite to modelling responses. By comparing two competing models of spike generation, we demonstrated that near-deterministic firing is compatible with measured higher-order statistics, while Poissonian spiking is not. We published these results in an international journal and presented them at an international conference.

The hierarchical model of visual processing we developed predicts the dependence of neural response correlations on stimulus content. We tested this prediction using recorded neural responses from the V1 of macaques, using our experimental design at the Ernst Strüngmann Institute in Frankfurt. We developed novel control procedures to deal with confounding effects between measured signals. The results confirmed the prediction of our model and corroborated the earlier finding that the secondary visual cortex is involved in the representation of textures. The results are published on a preprint server, are being submitted for publication to an international journal and have been presented at two international conferences.

The way stimuli are encoded in neural activity can be determined by decoding stimulus properties from recorded neural responses. We investigated whether second-order statistics play a significant role in the decodability of natural image stimuli from spike trains recorded from V1, and determined their importance. We also examined how decodability evolves over the course of an experimental session, and what aspects of the neural activity are essential to decode task-related experimental variables. We presented these results at two international conferences, and we are preparing them for publication in an international journal.

Using semantic memory to compress episodic memories. Continual learning, constructing and continually updating a model of a complex environment based on experiences, arose in recent years as a major field of interest in machine learning, while also being a longstanding challenge for cognitive science since human learning takes place in the same regime. One of the chief challenges in continual learning is how previously obtained information can be efficiently represented, that is, what kind of memory should a continual learning agent have. Traditional approaches based on optimising point estimates in deep learning architectures suffer from ‘catastrophic forgetting’: updating the model compromises performance on already learned tasks. While a Bayesian approach does not suffer from this issue, resource constraints that result in information loss render learning the structure of the environment impossible. In preceding years we have proposed that a combination of semantic and episodic memories can mitigate this issue and enable continual learning of model structure.

Our previous work contained the simplifying assumption that episodes are remembered verbatim, which is empirically known to be unrealistic for human learning and is a presumably wasteful use of memory resources. To remedy this, we have proposed that semantic memory, a latent variable hierarchical generative model of the environment, can be used to compress the episodes. We have argued that this corresponds to a specific choice of distortion function in rate-distortion theory, where predictive ability is prioritized over reconstruction of experiences in data space. We have shown that this choice explains robust biases in human memory errors in a classic experimental setting that tests memory for sketch drawings. To do this, we have approximated semantic memory for human sketches by training a latent variable generative model on the recently published QuickDraw database and shown that our compression algorithm qualitatively reproduces the distortions found in the experiments. We have shown on a simple model of hierarchical mixture of gaussians that the hierarchy of latent variables can enable compression with variable level of detail and approximate the optimal rate distortion curve. In the following year, we will test whether semantic compression of episodes can explain a greater variety of memory distortion experiments, and whether the inference of model structure is possible when episodes are compressed.

Cognitive tomography in an implicit learning task. Investigating human learning and decision making in dynamical environments in a general setting could allow one to understand the common principles relating intuitive physics, natural language understanding and theory of mind. Higher-level representations in temporal domains could then be measured for each individual.

We contributed to developing and improving methods for inferring human representations. To gather information in high-dimensional spaces, one requires a large number of data points during a learning process to identify the model forms individuals use during a learning task. The generative process of behavioural responses is, however, highly confounded with nonlearning- related effects. We developed a method for segregating the variation in response time measurements that are related to such confounds from the variation induced by learning. As a result of our analysis, we concluded that the confounds may impose a much larger effect on the response times than learning itself, rendering filtering or other form of accounting for confounds essential for inference. Our work was presented at two international conferences and is now in review at a journal for publication.

The next step in our research is to identify the cornerstones of efficient learning in a dynamical environment. There are a number of competing modelling ideas that can equally account for observable data, however, they impose qualitatively different inductive biases. Ideally, a normative account of the learning problem would point at specific sets of inductive biases, which in turn could be contrasted with human behaviour.

Another key to connecting learning of different temporal domains is the capability of extracting meaningful segments from a stream. This is equivalent to identifying events, actions, subgoals in the theory of mind domain, words, word-elements in the language domain, basic elements of motion, collisions and other types of interactions in the intuitive physics domain. The adequate discretisation of the stream is essentially tied to the temporal structure present in any given domain. Inference of the segmentation as well as of the structure over the segments is required to be handled simultaneously.

Public outreach activities. In our vision it is central to inform the public, and provide access to advances in the field of systems neuroscience and machine learning. We pursued three different paths to achieve these goals. In a number of media appearances, which included radio interviews, television shows, and written reports, we disseminated our research achievements. Traditional university lectures in a number of courses reached a large number of students with varying backgrounds arriving from a wide spectrum of universities. Finally, we have started a Junior Brain Computer Interface Lab for high school students. The goal of this lab is to provide access for a small team of high school students to the world of AI, machinel learning, and systems neuroscience through designing computational tools that can be used to control devices, such as small robots or computer games, on line by signals recorded from a so-called e-cog device.