Università degli Studi di Trento,
CIMeC, Centro Interdipartimentale Mente/Cervello, Rovereto

Istituto Italiano di Tecnologia,
CNCS, Center for Neuroscience and Cognitive Systems, Rovereto

A novel mathematical approach to the study of behavioral readout

PhD candidate: Monica Moroni
Advisor: Stefano Panzeri

Doctoral School in Cognitive and Brain Sciences

32nd Cycle

December, 13th 2019

ISTITUTO ITALIANO DI TECNOLOGIA
NEURAL COMPUTATION

CIMeC
Center for Mind/Brain Sciences

Abstract

Precise quantification of how spatio-temporal structures in neural activity are consequential for behavior, remains challenging. In this work I introduced a novel mathematical method to investigate behavioral readout. The method is based on the embedding of neural activity patterns into a metric space and on the definition of perceptual distances, that describe the internal representation of percepts. While being rigorously and mathematically well defined, the method is very general and flexible in the definition of perceptual distances. This makes it ideally suitable to the study of various systems, starting from single cells to neural populations. As a proof of principle, I applied the presented method to experimental data provided by Edmund Chong and Dmitry Rinberg at the Neuroscience Institute at New York University (NYU). We combined the novel method with an interventional approach to investigate the neural code used in the mice olfactory bulb (OB) to generate behavior. We trained animals to recognize artificial stimulation patterns on the OB and we systematically varied the stimulus features in order to thoroughly investigate the neural activity space. We were able to show some basic elements of OB coding: both spatial and temporal features of OB activity patterns are read out by animals, the readout favors early activity (primacy effect), spatial features are combined in an almost linear fashion, while temporal features are used in a more complex way. Our findings were replicated using logistic regression, a classical method, but our approach was advantageous in terms of model complexity and interpretability.

Acknowledgements

At the end of my doctoral studies, I would like to express my gratitude to those who supported me during these years.

First of all I would like to thank my advisor, Stefano Panzeri, for allowing me to collaborate with him and for guiding me through my research work. His guidance and suggestions have been fundamental for the design and fullfillment of this work. Together with Stefano, I would like to thank my colleagues for their useful advices during work hours (but also for the funny moments we shared).

I would like to thank also all the collaborators to this project and to the other projects I developed during my studies.

My special gratitude goes to Edmund Chong and Dmitry Rinberg from the Neuroscience Institute at New York University for the collection of the data used in this work and for the exciting conversations about the project. I cannot avoid to thank the funders of this project, that was funded by the NIH BRAIN INITIATIVE RO1 NIH 12590874 *Mapping of spatio-temporal code features to neural and perceptual spaces*.

I would like to thank also Andrea Sattin, Marco Brondi and Tommaso Fellin from the Optical Approaches to Brain Function Laboratory at IIT in Genova for allowing me to collaborate to other projects. For sake of time and clarity I couldn't describe discribe in detail in this thesis the projects I carried out with them, but I really enjoyed working at those projects.

Finally, I would like to express my gratitude to my family and my friends. Their support has been a key element to reach this goal. They were always there in difficult moments, to give me support and energy to face any kind of difficulty. And they were always ready to share with me the joy and the satisfaction for every small goal I achieved in these years.

Contents

List of Figures

Chapter 1

Introduction

Animals and humans are continuously interacting with the external world. In everyday life, they face different stimuli, record them through sensory organs, interpret the collected information and make choices according to this interpretation. For example, in the animal world, a prey must be able not only to see, hear or smell a predator that is approaching, but also to classify its presence as a potential danger and implement the best strategies in order to guarantee its own survival. This cognitive process of collecting, identifying, classifying and interpreting external stimuli is called *perception* and understanding its underlying mechanisms is one of the fundamental questions of neuroscience.

In order to be successful in perception, a system must be endowed with two fundamental properties: discrimination and generalization. Discrimination is the ability to distinguish different stimuli and lies at the core of perception. Recalling the previous example, this means that an animal must be able to distinguish a predator from any harmless animal. Given the importance of such property, a classical strategy to investigate perception consists in training subjects in discrimination tasks, where they have to report the match (or mismatch) between different stimuli, in order to understand which stimuli they can (or cannot) discriminate. On the other side, also the generalization property [61] is fundamental for perception. Such property consists in the ability to build abstract categories starting from single experiences. Again, going back to the previous example, a pray must run away from all the predators, not just

from a specific predator that it already met. The importance of the generalization property stems from the variability of the external world and of the internal representation of stimuli. Indeed, even in controlled experimental paradigms in research laboratories the same stimulus is rarely presented in exactly the same manner and it is well known that neural activity in response to stimuli is modulated from the current brain state, therefore producing different responses to the same stimulus. A system must be able to discard (or consider properly) these sources of variability when evaluating stimuli. A useful concept to build a bridge between the properties of discrimination and generalization is *similarity* (or dissimilarity). Similarity can be defined at different levels and therefore represents a tool to switch between discrimination and generalization. Among the many possible definitions of similarity, we want to highlight the concepts of *physical similarity* of a pair of stimuli and *perceptual similarity* of the same pair of stimuli. Given a set of physical measurable properties, physical similarity is defined as the overall quantitative difference in all the stimulus properties. It is fixed and neither depends on the observer nor on the kind of task. On the other side there is perceptual similarity, which is how different two stimuli are perceived. This depends on the physical properties of the stimuli, but is also influenced by the kind of task, by the context and by the subject experience, thus involving the generalization ability. For example, if two visual stimuli differ in shape and color and a subject is asked to rate their similarity based on shape, the effect of colors on perceptual distance will be small. Perceptual illusions (for example the Mcgurck effect [42]) provide an example that this distinction between physical and perceptual similarity is not an artifact, but effectively has some consequences on perception. While providing a unifying concept between discrimination and generalization, perceptual similarity is complex to investigate because it is an intrinsic measure of the subject percept and we can use only the subject reports to evaluate it. Usually, in order to address these issues, experimental paradigms attempt to control very precisely the stimulus and make the link between sensation and subject report as straightforward and transparent as possible.

The study of perception has been carried at different levels. Following a classification provided by Marr, analyses in neural systems can be studied at three levels: the computational (what problem does the system solve), the representational (what algorithms does it use to solve the

problem) and the physical level (what neuronal structures implement it) [52]. Historically, the first approach to the study of perception is *psychophysics*. Psychophysics was defined as "the analysis of perceptual processes by studying the effect on a subjects experience or behavior of systematically varying the properties of a stimulus along one or more physical dimension"[8]. Through psychophysics we can directly link stimuli to behavior and try to get insight into which physical features are important for discrimination and how sensitive subjects are to changes in those features. It has been shown that many different species perform discrimination of stimuli based on proportional differences insted of absolute magnitude. The law describing this proportional processing of stimuli magnitude is called Weber's law and is described formally as

$$\Delta I = kI \tag{1.1}$$

where I is the stimulus magnitude, ΔI is the minimum perceivable difference in stimulus magnitude and k is a constant value, called Weber fraction. Discrimination based on the Weber's law presents three features:

- the difference threshold in magnitude necessary to discriminate between stimuli depends on the magnitude of stimuli;
- the minimum perceivable difference increases when the stimulus magnitude gets larger;
- the error in stimuli discrimination is asymmetric. This means that if a subject is asked to detect a target stimulus and is presented with two non-target stimuli that have the same absolute difference in magnitude from the target, but in different direction, the error in the detection will be larger for the non-target stimulus with larger magnitude.

These findings are informative about the relevance of physical features of the stimuli and the way they inform behavior, but do not shed light into the mechanisms that neural systems implement to generate perception. In other words, the study of perception using psychophysics stops at the computational level (as defined by Marr). In order to go from the first level to the representational level, researchers integrated the investigation of perception with approaches complementary to psychophysics. One of these approaches, that will be described more in detail

in the next chapter, is computational modelling. In computational modelling, perception is the result of two main serial processes, that can be broadly described as 'sensation' of an external stimulus and 'decision making' or 'action' [22]. From an information processing point of view, these two main processes correspond to the following steps: *stimulus encoding* and *behavioral readout* of neural activity [50]. The separate study of these two processes started to unveil the brain mechanisms used to represent stimuli in the brain and to map neural activity to behavior. Recently, however, two main issues arose regarding the study of perceptual discrimination: the role of causality and correlation, and the need of a unifying framework for the investigation of stimulus encoding and behavioral readout. The first problem is due to our limits in the study of a complex system like the brain. When we study some functions, we focus our attention only on the recorded neural activity but this is highly modulated and correlated with the activity of not observed brain areas. It could happen that a feature is involved in stimulus processing according to observations and analyses, but this is due to its correlation with other not observed features which are causally linked to stimulus processing or behaviour. One possible way to overcome this issue became available with the development of interventional tools. Through silencing or artificially writing neural activity we can test not only correlational but also causal relationships, thus revealing the real components of neural codes. The need of a unifying framework, on the other side, stems from the observation that we are not guaranteed that the brain is an optimal observer and reader of neural activity. For example, information that is encoded with millisecond temporal precision or locked to some temporal onset can be used only if the brain is able to read neural activity with the same temporal resolution or has available information about the time the activity is locked to. For this reason, it is fundamental to look for neural features that simultaneously carry information about both stimulus and behaviour. The development of proper mathematical tools that measure the information at the intersection between encoding and readout will allow to unveil the neural code used for perception.

In this general framework of the study of perception, the work of this thesis fits in the study of behavioral readout. I will present a new approach to investigate the neural readout, that is based on a formal definition of the concept of perceptual similarity. First, I will present a review of the methods currently used to investigate perceptual judgments and behavioral

readout (Chapter 2). I will then introduce the mathematical concept of metric space, review its application in neuroscience and build a metric that maps the stimulus space to a perceptual space, representing the perceptual similarity between each possible pair of stimuli (Chapter 3). The metric can be then used to predict the discriminability between stimuli or the degree of generalization of the system across variation of stimuli. I will then show an application of this approach to artificially generated spatio-temporal patterns of neural activity, thought as artificial stimuli (Chapter 4). Using this approach we can investigate the sensitivity and specificity of neural readout, thus providing limits to the neural code used for perception. In the last part of this work, I will discuss the strength and the limits of the presented approach and present some possible future directions of the project (Chapter 5).

Chapter 2

Behavioral readout

As briefly illustrated in Chapter 1, perception is the result of many complex processes, starting from the reception and encoding of stimuli and resulting in some behavioral responses. A simplified view of perception describes it as a two-stage process. When a stimulus is presented, it elicits a response in the neural system, first in sensory receptors and later in downstream neural circuits. This is the first step and is called *stimulus encoding*. Information encoded about the stimulus is then processed, integrated and used in order to guide behavior. The process of converting neural activity into behavior is the second stage of perception and is called *behavioral readout*. In order to deepen the understanding of the processes that give rise to perception, researchers focused separately on these two steps generating perception.

2.1 Stimulus encoding

Stimulus encoding refers to the process of collecting and processing information about some features of the presented stimuli. Studying the encoding translates in asking which kind of information about stimuli is translated into neural activity and to which extent. The way in which the brain represents stimuli is called *neural coding*. Since neural activity consists of spatio-temporal patterns of activation of neurons, possible coding strategies can be based on neurons identity, timing of activation, synchronicity among neurons and other single neuron or

population features. Unveiling which of this coding mechanisms is implemented to represent stimuli features is the final aim of the study of the encoding process. This approach could be very helpful in setting an upper limit to the coding capacity of a neural system and in understanding which physical features are encoded by the brain. Lot of work has been done in the study of stimulus encoding and it has been shown, in different systems, that the brain is using both spatial and temporal features to encode stimuli, that the neural code is a population code, meaning that the joint activity of multiple neurons carries more information about stimuli than the activity of single neurons. At the same time some redundancy is present and also correlations play a role in the neural code. While providing important insights into brain mechanisms used to represent sensory information, the study of only the encoding process does not fully answer the question about the mechanisms underlying perception. Indeed, the encoding process is completely uncorrelated from behavior: only stimuli and neural activity are used for this kind of investigation and a behavioral report from the subject is neither necessary nor required. Rephrasing the concept, we could say that studying the encoding process answers the question of what an observer could say about the stimulus just looking at the neural activity. Furthermore, there are two main limits of the study of stimulus encoding by itself: first of all, we are not guaranteed that the brain has an optimal downstream readout of neural activity. For this reason, if some stimulus features are encoded in a way that cannot be read out, they cannot play a role in perception. In second place, it is known that perception is modulated by other variables (on the top of the stimulus), such as attention, learning, motivation, reward/failure rate feedback and so on. Therefore, studying the encoding of stimuli without a behavioral task might lead to inaccurate conclusions about perception.

2.2 Behavioral readout

Behavioral readout refers to the process of translating neural activity into a choice or a behavioral response. This is a fundamental part of the perception process and the study of this second stage of perception, as we have seen in the previous section, cannot be ignored. Studying the behavioral readout consists in investigating which neural features carry information about

the choice and aims at answering questions about which neural features are relevant for behavior and how sensitive systems are to variations of those features. In contrast to stimulus encoding, this process is not independent from behavior, and researchers need data that contain both the neural activity (that is read out by the system to produce behavior) and a behavioral outcome. From an experimental point of view this is more complex than recording activity without a behavioral response, because subjects must be awake and trained to perform the designed tasks. For this reason, behavioral readout has been studied less and still needs further investigation. Obviously, the study of the single readout process is not sufficient to investigate perception: this process is independent from the stimulus and we don't know whether the same features that are read out contain information about the stimulus. In other words, investigating the readout answers the question of what an observer could say about behavior just looking at neural activity (without any information about the stimuli). In the following part of this chapter, I will provide an overview of some commonly used methods to investigate the role of neural variables in perceptual readout.

2.2.1 Choice Probabilities (CPs)

A classical approach for studying percept formation in brain regions consists is the computation of choice probabilities (CPs), introduced by Britten et al in [7] and reviewed in [14]. CP is a measure of the relation between trial-to-trial fluctuations in the firing rate of a neuron and subject's behavior and can be interpreted as the accuracy of an observer of neural activity in predicting the animal choice. Given neural responses and behavioral choices, CP uses ROC (Receiver Operating Characteristic) analyses to compute the discriminability of the neural responses distribution for the different choices for a fixed stimulus. The stimulus is usually selected such that it is close to the threshold for discriminability. If CPs do not strongly vary as a stimulus function, some authors compute a grand CP by standardizing responses across stimuli and computing a single value for CP. A value of CP close to 1 indicates that neural activity is different for different choices, therefore neurons represent an internal percept. A value of CP close to 0.5 indicates that there is no detectable relation between neural activity

and behavior. Since CP is not related to the stimulus but only to behavior, it can be considered as an internal track of percept. Since its introduction, CP has been widely used in order to investigate the role of neurons in forming percepts in different brain areas [7, 15, 26, 46, 72, 76]. A common result is that CP increases from early sensory areas to higher order areas, while the stimulus representation decreases along the same pathway. This is in accordance with the idea that the brain processes progressively transform the identity of the stimulus into a more abstract percept, that is then used to generate behavior. However, while being significantly different from chance, mean values of CP in many brain areas are usually low (CP ~ 0.55 in [7], CP ~ 0.54 in [46]). This could be expected when using stimuli that are not encoded from the analysed neurons, but is at odd with the fact that some neurons contain high values of information about stimuli. A possible explanation for this observation is that CP is variable across neurons and mean values might not be representative [7]. Another reason for low values of CP might be the temporal window used to compute this quantity [76]. It is possible that the generation of percept follows the stimulus onset with some delay and averaging CP in a longer temporal window might not be very informative. Indeed, it has been shown that peak values of CP can be much higher than the mean value. A third possible explanation could derive from the downstream steps of neural processing [60]. If the activity of many independent neurons is pooled together, than low values of CP might combine to generate a reliable percept. However, this is at odd with results showing that the activity of neurons is correlated and information from many neurons is redundant. While some works claim that CP is a measure that shows which neural features affect behavior, there is evidence that it is not a causal measure and that also neurons which are not used for the readout can have CP values larger than chance level. In particular, using a model based on pools of neurons to simulate the mechanisms generating CP, Shadlen et al. showed in [60] that CP is affected by noise correlations among neurons and by the correlations structure. Indeed, if any neuron is correlated with another neuron causally involved in perception, due to correlations its value of CP will be larger than 0.5.

It is evident, therefore, that CP presents lots of caveats that should be considered when performing analyses. A first limit is due to the need of a sufficient number of trials (at least 100 trials from [7]) in order to have a reliable estimate of CP and, on top of this, the trials

should contain both choices, making impossible to compute CP for strongly tuned stimuli. Another caveat of CP, maybe among the most relevant for its interpretation, is that it is a correlational and not a causal measure. For this reason, neurons with CP larger than chance might not causally intervene in the percept formation, but they might be simply correlated with causally relevant neurons. Indeed, from modelling approaches correlations appear to have the strongest impact on CP. Based on the same model of [60], in [25] Hefner et al. derived a mathematical formulation for CP starting from neurons correlations and readout weights. It emerged that CP is a function of both readout weights and correlations and it cannot be therefore used to assess the causal role of neurons for behaviour. Another caveat of CP is that we do not know with certainty whether it derives from feedforward mechanisms or if it reflects a top-down modulation of neural activity. Some authors [7] suggest that the time course of CP provides evidence of its generation from a feedforward mechanism, while others suggest that other mechanisms might be involved. For example, Hernandez et al show in [26] that CP in higher order brain areas is not significantly delayed with respect to CP in earlier areas which is in contrast with a feedforward mechanism, while Nienborg and Cumming in [46] claim that CP comes from a top-down modulation, by showing that sensory relevant cues and decision-related neural activity time course are not correlated. Since it is well known that neural code is a population code and noise correlations appear to be very important in computing CP, different approaches have been applied to investigate the role of correlations in behavioral readout. More in details, some authors suggested a correlation between decreased noise correlations and improved behavioral readout [13], but the directionality of the relationship was not investigated. In some cases it was proposed that the temporal structure of non-causal measures such as CP was a hint for the exclusion of top-down mechanisms [45]. More complex approaches, based on single-trial measures, have been proposed as well [45]. Opposite evidence was provided in different studies [21] showing that decorrelation in neural activity, even though correlated to improvement in performance, does not play a relevant role in the readout. The existence of controversial interpretation of CP and similar methods pushes for the application of new experimental or analytical tools that might allow to unveil the causal role of neural features.

2.2.2 Generalized Linear Models (GLMs)

A different approach to the study of behavioral readout consists in applying generalized linear models (GLM) to predict behavior. In "canonical" linear models, a response variable Y is modelled as a weighted linear combination of some predictors $X_1, X_2, ..., X_N$ (or regressors).

$$Y \sim \beta_0 + \sum_{j=1}^{N} \beta_j * X_j = \beta_0 + \boldsymbol{\beta^T X} \tag{2.1}$$

If we denote with $\{y^1, y^2, ..., y^n\}$ the realizations of the variable Y and with $\{x_j^1, x_j^2, ..., x_j^n\}$ the realizations of the regressor X_j, the formulation of the linear model is the following:

$$y^i = \beta_0 + \sum_{j=1}^{N} \beta_j * x_j^i + \epsilon^i = \beta_0 + \boldsymbol{\beta^T x^i} + \epsilon^i \tag{2.2}$$

where β_0 is the *bias term*, β_j are the weights of the predictors and ϵ^i is the *error term*, usually drawn from a normal distribution.

Such models are very powerful in describing data, but there are some cases where they cannot be applied, for example when the response variable takes values only in a limited range. GLMs have been introduced to overcome some of the limitations of the linear regression models. In GLMs each observation of the dependent variable Y is assumed to be generated by a distribution in the exponential family ($y^i \sim Y^i$). The mean value of such distribution $\mu^i = E[Y^i]$ depends on a linear combination of the predictor variables through a *link function* g in the following way:

$$E[Y^i] = \mu^i = g^{-1}(\beta_0 + \sum_{j=1}^{N} \beta_j * X_j^i) = g^{-1}(\beta_0 + \boldsymbol{\beta^T X^i}) \tag{2.3}$$

When the response variable Y is a binary variable (assuming only value in $[0, 1]$), a common choice for the distribution for Y is the Bernoulli distribution. The mean value of Y^i can then be interpreted as the probability p of Y^i to assume a value of 1. A typical link function in this case is the *logit* link (or logistic function):

$$g(p) = ln(\frac{p}{1-p}) \tag{2.4}$$

Logistic regression is very suitable to behavioral data, which often consider as response variable the outcome of a binary choice. For this reason, in the following, I will focus on logistic regression. The parameters β_i of logistic regression models can be fitted using maximum likelihood estimation, that consists in maximizing the so-called maximum likelihood function or minimizing the negative log-likelihood function:

$$\min_{\beta_0,\boldsymbol{\beta}} l(\beta_0, \boldsymbol{\beta}, \boldsymbol{x}, \boldsymbol{y}) = \min_{\beta_0,\boldsymbol{\beta}} [-log(L(\beta_0, \boldsymbol{\beta}, \boldsymbol{x}, \boldsymbol{y}))] = \min_{\beta_0,\boldsymbol{\beta}} \{-\sum_{i=1}^{n} [log(P(y^i|\boldsymbol{x^i}, \beta_0, \boldsymbol{\beta}))]\} \tag{2.5}$$

where $P(y^i|\boldsymbol{x^i}, \beta_0, \boldsymbol{\beta})$ is the probability of observing the realization y^i conditioned on the observations $\boldsymbol{x^i}$ and the parameters β_0 and $\boldsymbol{\beta}$. Unlike for linear logistic regression, the maximum likelihood estimate is not guaranteed to converge to a solution. This happens especially when a large number of selected predictors have null weights or the predictors variables are highly correlated (multicollinearity) and fitting strategies are not able to assign uniquely to each predictor its effects. A solution to this problem was introduced by *penalized logistic regression.*

In penalized logistic regression a penalty term is incorporated into the maximum likelihood function in order to have a more stable and accurate estimate of the model parameters, especially in case of high dimensional or correlated data. The penalty term can assume different forms. In the following I present the minimization problem for *elastic net regression*:

$$\min_{\beta_0,\boldsymbol{\beta}} \{l(\beta_0, \boldsymbol{\beta}, \boldsymbol{x}, \boldsymbol{y}) + \lambda[(1-\alpha)\|\boldsymbol{\beta}\|_2^2/2 + \alpha\|\boldsymbol{\beta}\|_1]\} \tag{2.6}$$

where the function $l(\beta_0, \boldsymbol{\beta}, \boldsymbol{x}, \boldsymbol{y})$ is defined in Equation 2.5. The parameter λ is a regularization parameter that defines the strength of the regularization. When $\lambda \to 0$ the method converges to the maximum likelihood estimation, while for large values of λ the penalty strongly affects the parameters estimate. The parameter α controls the elastic net penalty and determines the type of contraints on the predictors coefficients. For $\alpha = 1$ the approach is equivalent to the Lasso regularization [70], which reduces the number of predictors of the model setting their weights to 0. For $\alpha = 0$ the model becomes a ridge regularization model [27], which reduces the complexity of a model by shrinking the values of the predictors weights.

Application of GLM models to neuroscience is becoming more and more widespread. These models allow to consider the effect on behavior of many different features and to discount for the effect of those features researchers are not interested in, such as non-neural variables (running speed and similar), therefore providing a tool to understand which neural features can explain the animal behavior [9]. Considering more than a single neural feature to predict behavior is a significant enhancement with respect to CPs. Using GLMs to model the data, Katz et al. replicated the finding (obtained using CPs) that information about behavior increases progressively from sensory areas to higher order areas [44]. Other studies used GLMs to show that posterior parietal cortex (PPC) has a functional role in sensory perception (both visual [39] and auditory [1]). This was not possible using CP, that allows to investigate only a single feature at a time.

2.2.3 Discussion

It is evident that the question about how perceptual representation is constructed and processed in neural systems is still open. Past research mostly focused on characterizing the perceptual information contained in single neurons. This provided a useful tool to link neural activity to behavior and to start understanding the neurophysiology underlying perception. However, it is well known that the neural system uses spatio-temporal patterns of activity to represent information. Studying single neurons properties, even though very helpful in getting insights about neural mechanisms, is not sufficient to characterize the neural code. Only recent studies focused their attention on population properties, such as correlations or global modulation [47], and on their role in perception, but at this moment results are inconsistent across studies. A second, maybe even more important limit of past studies is the lack of causality measurements. The methods presented in the previous sections, CPs and GLMs, provide a measure of the correlations between neural features and behavioral responses. They have been applied to describe behavioral responses to external stimuli as a function of the available neural measurements. However, in the last years, a criticality of these approaches has emerged: statistical results are very useful in revealing correlational links between neural

activity and behavior, but cannot be used to infer a causal (or functional) role of neural activity for perception [50]. Indeed, actvity in the brain is highly correlated and a simple description of behavior as function of some features does not support evidence for the active role of those features in perception. In order to infer something about causality, an interventional approach is necessary.

Thanks to the continuous development of new experimental and interventional techniques, researchers have now at their disposal a number of tools that open the way for alternative approaches to the study of behavioral readout. The manipulation of neural activity can now be done in many different ways. A first technique is pharmacological intervention, that can be used to block or enhance neural activity. Combining pharmacological inactivation of neural activity and the use of GLMs, Katz et al. in [36] provided evidence that neurons in LIP, while being correlated with behavior, are not causally involved in the decision-making process. However, pharmacological treatments do not have neither a good spatial nor a good temporal resolution and neural circuits might be able to reorganize in order to face the changes induced by this kind of intervention. For this reason, results obtained through pharmacological intervention must be interpreted very carefully. A second approach consists in electrical stimulation [69, 78, 57, 79]. Electrical stimulation can be very precise in time, but even this approach suffers from the lack of spatial resolution.

A revolutionary contribution to neural systems perturbation was provided by the development of chemogenetic and optogenetic tools to manipulate neural activity. Chemogenetic manipulate type-specific neurons through the interactions with small molecules injected in tissues. While lacking the single cell resolution, chemogenetics is suitable for long timescale experiments because its effects persist for several hours [68]. Optogenetic uses light to selectively control the activity of neurons. The spatial specificity of optogenetics ranges from controlling full populations of specific neuron classes to selectively controlling identified neurons with near millisecond precision [48]. Optogenetic intervention can be both excitatory, triggering target neurons firing, or inhibitory, silencing target neurons [17]. Furthermore, optogenetic is robust and can be used to manipulate neurons even in experimental sessions that take place in different days. A disadvantage of optogenetic is that it is power consuming and can not be

used for very long experiments, because continuous stimulation could damage neurons. Using optogenetic disruption of neural activity, Licata et al. in [39] claimed that posterior parietal cortex (PPC) has a causal role in the processing of visual stimuli and not auditory stimuli. Recent developments and improvements in the use of optogenetics allow to think about new approaches, where external stimuli are replaced by artificial generation of neural activity. Given the complexity of both spontaneous and stimulus-evoked neural activity, this approach could be helpful in disentangling the role of different neural features for behavioral readout. Indeed, the chance of stimulating directly specific neurons with single cell and single spike precision would make it possible to decorrelate neural features that are strongly correlated in the naturally evoked neural activity patterns and to study their role separately. An obvious drawback of such approach is the fact that artificially generated neural activity patterns are not necessary similar to the ones evoked by stimuli. Depending on the experimental question, such limit could be overcome by building up a "library" of neural responses to naturalistic stimuli (recording neural activity during the stimuli presentation) and selecting the articifial patterns from this library.

Chapter 3

Metric on neural space

In this chapter, I introduce and develop the idea of defining metrics on the space of neural activity to study perception. First of all, I provide the mathematical definition of metric space. Then, I give an overview of the main results obtained defining neural metrics for the study of stimulus encoding. Finally, I describe a possible application of metric spaces in the study of behavioral readout.

3.1 Mathematical definition of metric space

Before describing the use of metrics in neuroscience and discussing the advantages of such an approach, I present a formal mathematical definition of the concept of metric space. Given a set of elements $S = \{S_1, S_2, ...\}$, which later could be the set of stimuli or the set of neural activity patterns, a metric D is a mapping function (commonly called *distance*) that assigns to each pair of elements (S_i, S_j) in S a real number satisfying the following conditions:

- Non-negativity: $D(S_i, S_j) \geq 0, \quad \forall S_i, S_j \in S$. The distance between two elements must always be non-negative.
- Identity of indiscernibles: $D(S_i, S_j) = 0 \iff S_i = S_j, \quad \forall S_i, S_j \in S$. The distance between two elements is zero only in the trivial case where the elements are the same.

- Symmetry: $D(S_i, S_j) = D(S_j, S_i), \quad \forall S_i, S_j \in S$. Changing the order of the terms the distance does not change.

- Triangle inequality: $D(S_i, S_k) \leq D(S_i, S_j) + D(S_j, S_k), \quad \forall S_i, S_j, S_k \in S$. The distance between two points in the space cannot be larger than the sum of the distance between the first point and any other third point and the distance between the same third point and the second point.

The set S equipped with the metric D is called *metric space*. In case the second condition is not satisfied, the space is defined *pseudometric space*. It is always possible to transform a pseudometric space into a metric space by defining *equivalence classes of elements* as the sets of elements with null distance. The space of the equivalence classes of elements equipped with the original distance is a proper metric space. If the metric defined on a space is the classical Euclidean distance, then the space is called *Euclidean space*.

3.2 Metric space to study stimulus encoding

The introduction of metric spaces in studying neural activity dates back to the activity of Victor and Purpura ([74, 75]). The aim of their work was to investigate the nature of temporal coding by finding a relationship between natural stimuli and temporal patterns of neural activity. Classical ways of investigating stimulus encoding consisted in binning spike trains in bins of a given width and comparing the binned activity in response to different stimuli. However, such an approach presents the drawback that the dimension of the space becomes very high when the temporal resolution is very fine (short time bins) and it relies on the assumption that spike trains can be embedded in a vector-space, which was called into question in other works [28]. For this reason, Victor and Purpura suggested to build a minimal structure to describe spike trains (dis-)similarity, without using a vector space with Euclidean properties. They considered spike trains as points in an abstract space and introduced a definition of distance between pairs of points on this space. The main advantage of such an approach is that it is formally well defined, but at the same time allows for great flexibility, because the distance

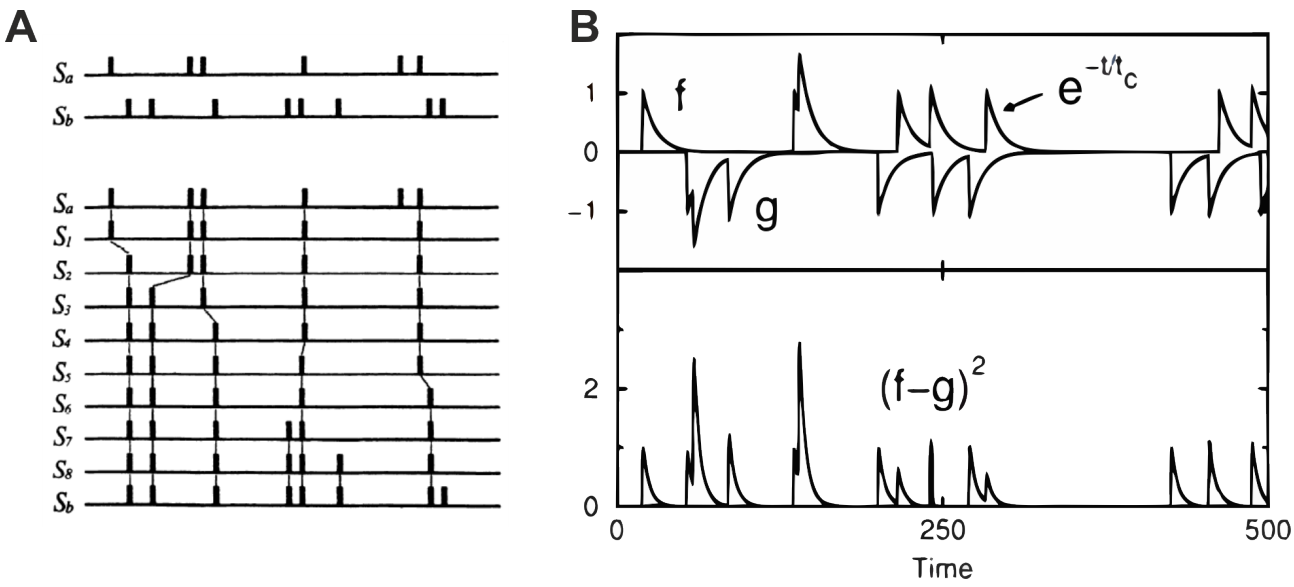

Figure 3.1: **Victor-Purpura and van Rossum distance. A)** The Victor-Purpura metric. Top. Two example spike trains S_a and S_b. The distance between spike trains is given by the minimum cost necessary to transform S_a in S_b using few allowed transformations. Bottom: the path with the elementary steps necessary to the transformation (from [74]). **B**) The van Rossum distance. The spike trains are transformed into continuous signals (one flipped) through convolution with an exponential kernel and their squared distance is the area under the squared difference (from [53]).

can be defined in an arbitrary way, provided that metric properties are satisfied. Indeed, the authors presented two different definitions of distance, based on two different intuitions about the possible biological mechanisms underlying neural coding. The first measure, D^{spike}, shown in Figure 3.1(A), is sensitive to the exact timing of spikes and relies on the idea of neurons as coincidence detectors. The second measure, $D^{interval}$, is sensitive to changes in the inter-spike interval and is related to the idea that the relevance of a spike can depend on the time interval from the preceding spike (because of mechanisms as long-term potentiation/depression). Both measures depend on a parameter, q, that represents the temporal resolution of the metric. For $q = 0$ the two metrics converge to the spike count metric D^{count}, where only the number of spikes is used to evaluate the similarity between spike trains. The two distances provide two different geometries in the spike trains space: two spike trains can be closer in the space equipped with one distance than in the space equipped with the other. If the features used to define a metric correspond to the features used to encode information about stimuli, we expect spike trains corresponding to the same stimulus to be close in the metric space. The authors implemented a clustering approach to test, for different values of the parameter q,

which metric better corresponds to the stimulus space and they measured the goodness of the stimulus-dependent clustering using an information-theoretical measure. They observed that both spike timing and inter-spike interval contain more information about stimuli than the simple spike count and, by tuning the parameter q, they were able to estimate the temporal sensitivity of the code used to represent stimuli.

Starting from the Victor-Purpura spike train distance, other metrics have been developed to measure the dissimilarity of spike trains. One of the most common is the van Rossum metric [53], shown in Figure 3.1(B). The main difference from the Victor-Purpura metric is that spike trains are transformed into continuous signals, before computing their difference, through convolution with an exponential kernel. The choice of the exponential kernel (instead of other kernels such as a Gaussian kernel or a square pulse) is due to its causal properties and to its biological plausibility. The distance between two spike trains is then defined as the Euclidean distance between the continuous signals corresponding to the spike trains. While reintroducing Euclidean properties to the metric space, such definition is easier to compute and less prone to ambiguities in case, for example, of different number of spikes in two spike trains.

Both the Victor-Purpura distance and the van Rossum distance are defined on single spike trains. However, with the development of more advanced recording techniques and the availability of multineuron data, the need for more sophisticated tools, which describe activity at a population level, emerged. The notion of metric was therefore extended in order to include both spatial and temporal features [3]. The new metric was characterized by two parameters: q, that represents the temporal sensitivity of the system as in the original Victor-Purpura distance and k, that represents the sensitivity of the code to neurons identity. The parameter k ranges between 0 and 2 and provides a gradual shift from a *labelled line* code (LL, $k = 2$), where each neuron activity is considered independently and cannot be interchanged with the activity of other neurons, and a *summed population* code (SP, $k = 0$), where only spike timing is used without regards for the identity of spiking neurons. The authors applied this metric approach to experimental data and observed, repeating the same analyses as [74], that both temporal and spatial coding is used to encode stimuli. In a similar way, also the van Rossum metric was extended to the multineuron case [29].

3.3 Metric space to study behavioral readout

In the previous paragraph, we showed that the introduction of metric spaces in the investigation of stimulus encoding contributed to important progresses in the understanding of the neural code. In particular, it revealed possible biologically plausible coding strategies and allowed to make some inference about the temporal and spatial sensitivity of the neural code. Given the success in applying such approach to the investigation of stimulus encoding, we queried whether the same or a similar approach would be successful also in the study of perceptual readout. The general idea consists in defining a metric on the neural activity space that maps each pair of neural activity patterns into a non-negative quantity that represents the *perceptual distance* between the considered neural patterns. This means that the perceptual metric should reflect similarities in the elicited perceptual reports and not similarities in the neural patterns structure. In other words, we want to build a function such that distances between patterns are small when subjects are not able to discriminate between them and increase as patterns get more and more distinguishable. Since we do not have a measure of the internal percept, we must rely on perceptual reports (a binary choice in many task designs) in order to evaluate how precisely the metric reflects perceptual readout. We suggest to use a link function, such as the logistic function (see Eq. 2.4), to map the computed distances into behavioral reports and then to compare the behavioral data with the obtained predictions. We expect that predicted behavior will match behavioral data in case the neural features used to build the metric function are the same that are read out by the system to inform behavior. We will call the space we consider, endowed with the metric, *perceptual space*. As for spike train distances, we will try to define the metric making assumptions on plausible underlying computational mechanisms. However, we do not claim that the computations used to define the metric are those biologically implemented by the brain. From an experimental point of view, in order to implement this approach we need experiments where we both record neural activity and track perceptual outcome. Thanks to the development of techniques to record neural population activity in awake animals, such data are now available and allow the test of the suggested approach.

Given the countless ways of defining metrics, the presented approach is generalizable to

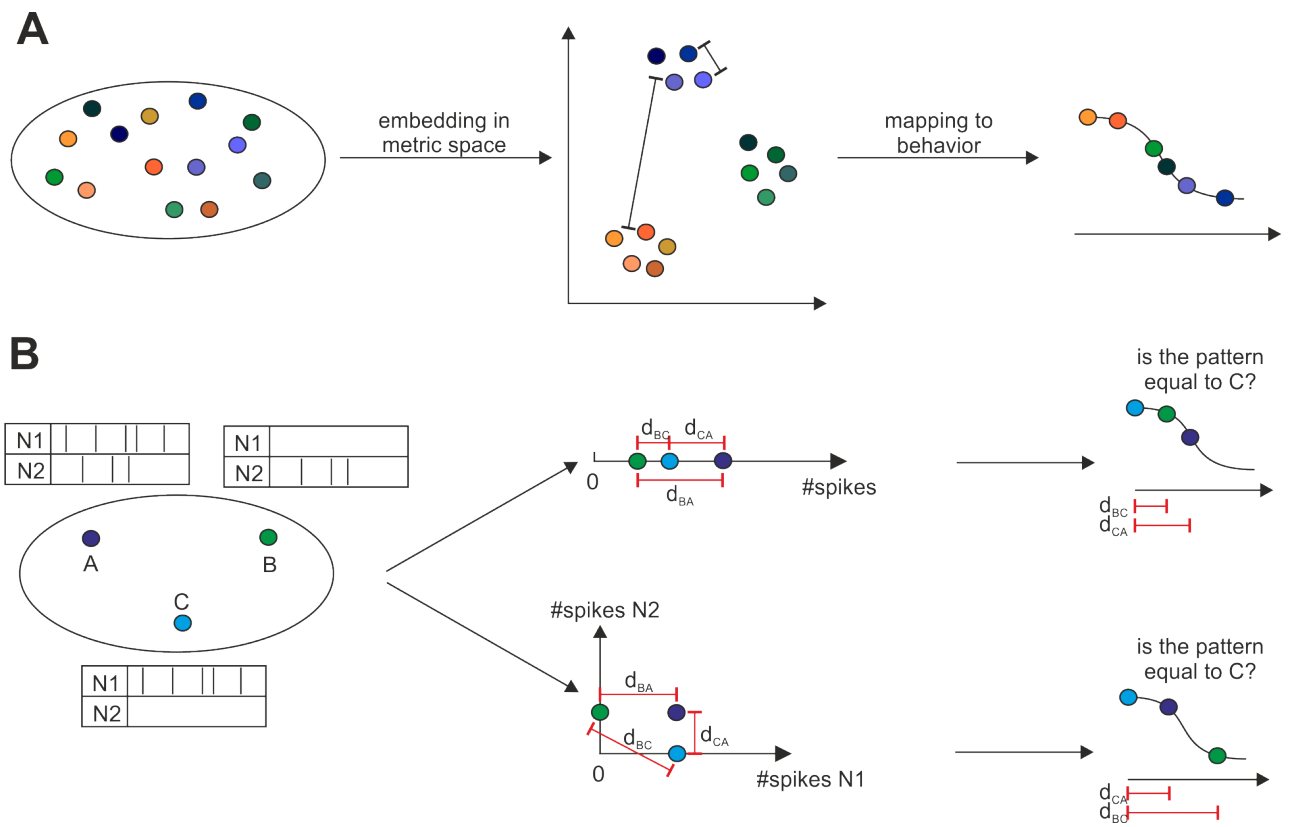

Figure 3.2: **General illustration of the perceptual metric model. A.** General illustration of the metric approach to study behavioral readout. Patterns of neural activity (represented as colored dots) are first embedded into a metric space and then mapped into behavior using a link function (in this example a logistic function). **B.** An example of two simple metrics on the space composed by spike trains of two neurons. The metric definined at the top is based on the summed spikes count, while the metric defined at the bottom is based on the spikes count of the single neurons. As shown in the logistic curves on the right, different metrics lead to different behavioral outcomes. In the metric defined at the top, the probability of reporting a match with pattern C is higher for pattern B than for pattern A. In the metric defined at the bottom, instead, the opposite prediction holds.

the study of almost any neural system. Theoretically, we can define distances using any computational process we might think is implemented by the neural system, under the constraints that the metric properties are fullfilled. As GLMs, this approach allows to consider on a trial-to-trial basis the effect of many features on behavior but metrics are more general than GLMs because they could model also non linear relationships.

To our knowledge such an approach has been poorly investigated and applied to the study of perceptual representation [58]. An example of a perceptual metric can be found in [71], where the authors derive a retinal metric to quantify the perceptual similarity of visual stimuli. While providing a way to define metrics on neural space, our approach is intrinsically different from

the approach in [71], due to how the metric is built. Indeed, we propose to process neural patterns using biologically plausible computations in order to build a metric. The predictions of the metric are then compared to behavioral outcomes to evaluate whether the metric reflects coding mechanisms. The authors in [71] adopt a reverse approach: they define the metric using behavioral outcomes and then infer brain computation processes and neural coding mechanisms from how neural patterns are distributed accordingly to the obtained metric.

Chapter 4

A perceptual metric on olfactory bulb spatio-temporal neural patterns

4.1 Introduction

A familiar object evokes a complex pattern of activity in the brain, but it is possible that only a structured subset of this activity, representing critical combinations of sensory attributes, is essential for recognition. A key challenge is to identify the subspace of neural activity that induces the percept. This activity may consist of multiple spatial or temporal features, such as which cells respond and when they respond, relative to stimulus onset or each other. Do individual features contribute differentially to the formation of the percept? For example, the activity of some cells in a pattern may be more important than others. Does the formation of the percept depend on how features are combined? The sequential activation of multiple cells or the latency of their activation relative to brain rhythms are examples of feature combinations that may be perceptually meaningful [10, 54].

The difficulty in addressing these questions is two-fold. First, multiple features co-vary with perceptual responses, making it difficult to disentangle their independent contributions to perception. Previous studies have mainly focused on correlating neural activity with perception, where the contributions of individual features are entangled. Second, we lack a single framework

for quantitatively comparing the perceptual contributions of individual and combined features. For example, we know from causal manipulation studies that single neurons [30] or small timing differences [65, 64, 78] can affect perception, but not their relative importance and how they come together in larger patterns to produce perception.

Here we developed a novel framework for finding the perceptually-meaningful spatiotemporal subspace of neural activity. We optogenetically manipulated individual activity features independently of other features, manipulated combinations of features, and compared the effects of all manipulations under a common metric.

We used mouse olfaction as our model system to study these questions, as odor perception is correlated with complex spatio-temporal activity patterns, and these patterns can be optogenetically manipulated in mice while measuring their perceptual responses. We chose such system for two main reasons. First, it is simple, which means that contains few processing stages of sensory information. Second, it is manipulable, which means that experimenters have tools to intervene directly on neural activity. These two features, combined together, make such system suitable for a causal investigation of behavioral readout.

When an odor is presented to a mouse, chemicals composing the odors enter the animal's nose and activate olfactory sensory neurons (OSNs) collocated in the olfactory epithelium. Each OSN is responsive to specific chemicals and multiple OSNs responding to the same chemical maps to a glomerulus in the olfactory bulb (OB) [6]. From the OB the activity is transmitted to mitral/tufted (M/T) cells and from there relayed to the olfactory cortex. On the top of this forward circuit, there is interglomerular inhibition and a feedback inhibitory signal coming back from the cortex to the OB. Since different chemicals activate different OSNs, different odors result in the activation of different glomeruli in the OB producing characteristic spatio-temporal patterns of neural activity.

Different hypotheses have been formulated about odor coding. A first hypothesis claims that mice use a spatial (or identity) coding to represent odors identity and concentration [37]. A spatial code would be quite simple to implement but, at the same time, very powerful. Indeed, in mice each OB contains ~ 1800 glomeruli [43]. If different combinations of activated glomeruli

can be read out and discriminated, the coding capacity of the system becomes extremely large and, in principle, such a combinatorial code would be sufficient for odor coding. However, it has been shown that chemicals concentration affects not only the identity of the activated glomeruli, but also the latency of neural responses [67]. If the brain is capable of reading out also activation timing, temporal coding would be another plausible candidate for odors concentration encoding. A few studies already focused on the readout of OB activity, that is the mapping of OB spatio-temporal patterns to behavior. Applying mathematical models to simulated data or odor evoked responses, it has been shown that both spatial and temporal features are used to inform behavior [66, 41]. However, studying the neural code using natural stimuli can be difficult because naturally evoked neural activity patterns present correlated spatial and temporal features and do not allow to investigate spatial and temporal coding independently. This could lead to spurious or ambiguous conclusions about coding mechanisms.

A way to avoid correlated data, probe causal relationships in neural systems and disentangle between ambiguous conclusions consists in applying of an interventional approach [50, 33]. Optogenetics studies already revealed that the mouse olfactory readout is sentitive to single glomerulus changes in activation timing and intensity [64] and lead to the hypothesis of a primacy code for odor discriminations, which means that only OB activity at the beginning of a sniff cycle is functionally relevant for behavior [77].

Leveraging on such interventional approach, we wanted to extend the study of behavioral readout from the single glomerulus level to the population level. To achieve this, we generated synthetic odors by directly activating multiple OB spots through optogenetics and manipulating independently the identity of the activated spots and their activation timing. In this way we aimed at systematically exploring the space of OB neural activity in order to understand the distinct role of identity and timing in the neural code and to characterize the sensitivity of the system to perturbation of single features. Hence, by performing precise and parametric manipulation of glomerular activity and quantifying effects under a common metric, we derived a unifying model that explains how odor perception arises from structured glomerular activation.

4.2 Methods

4.2.1 Experimental design

We chronically implanted OMP-ChR2-YFP mice (n=6) [65] with cranial windows to expose dorsal OB, and performed optogenetic stimulation using a digital micromirror device system. Spatio-temporal stimulation patterns were projected onto the OB of the head-fixed mice in front of lick spouts and a pressure sensor for respiration monitoring (Figure 4.1A). We first characterized single spot stimulations: similar to previous reports [24, 16] and consistent with known anatomy, single spot stimulation (120x120 μm^2, 80 ms duration, 15 mW/mm^2) activated mitral/tufted (MT) cells, leading to instantaneous MT cell firing rates up to $\sim$100 Hz, with excitatory responses lasting around $\sim$80 ms, comparable to odor-evoked responses [63]. We verified that spots at the same stimulation parameters were perceptually detectable, but only for ChR2-positive mice, and without systematic spatial biases. We then used the same basic stimulation parameters for the main experiments.

Mice were trained in a 2-alternative forced choice task, in which 'left lick' and 'right lick' were randomly assigned to Target and Non-target patterns for each animal. Target patterns comprised of six spots, initialized randomly but fixed across subsequent sessions, activated in an ordered sequence defined in time where $t = 0$ marks inhalation onset. Non-target patterns were six off-Target spots, randomly chosen from trial to trial, with randomized timing within 300 ms from inhalation ($\sim$ single sniff) (Figure 4.1B). Each animal was first trained to discriminate between one Target and one Non-target. When the animal reached criterion performance of 80%, it was trained to discriminate between one Target and many randomly initialized Non-target patterns. In test sessions, probe trials were introduced. Probe trials accounted for 10% of the trials in each session, while Target and Non-target were 45% of the total number of trials each. Probe trials consisted in perturbations of the Target pattern were the identity and/or the timing of activation of the Target spots were changed. We considered four types of Probe trials:

- *Spatial perturbations*: in these trials one or more Target spots were replaced with

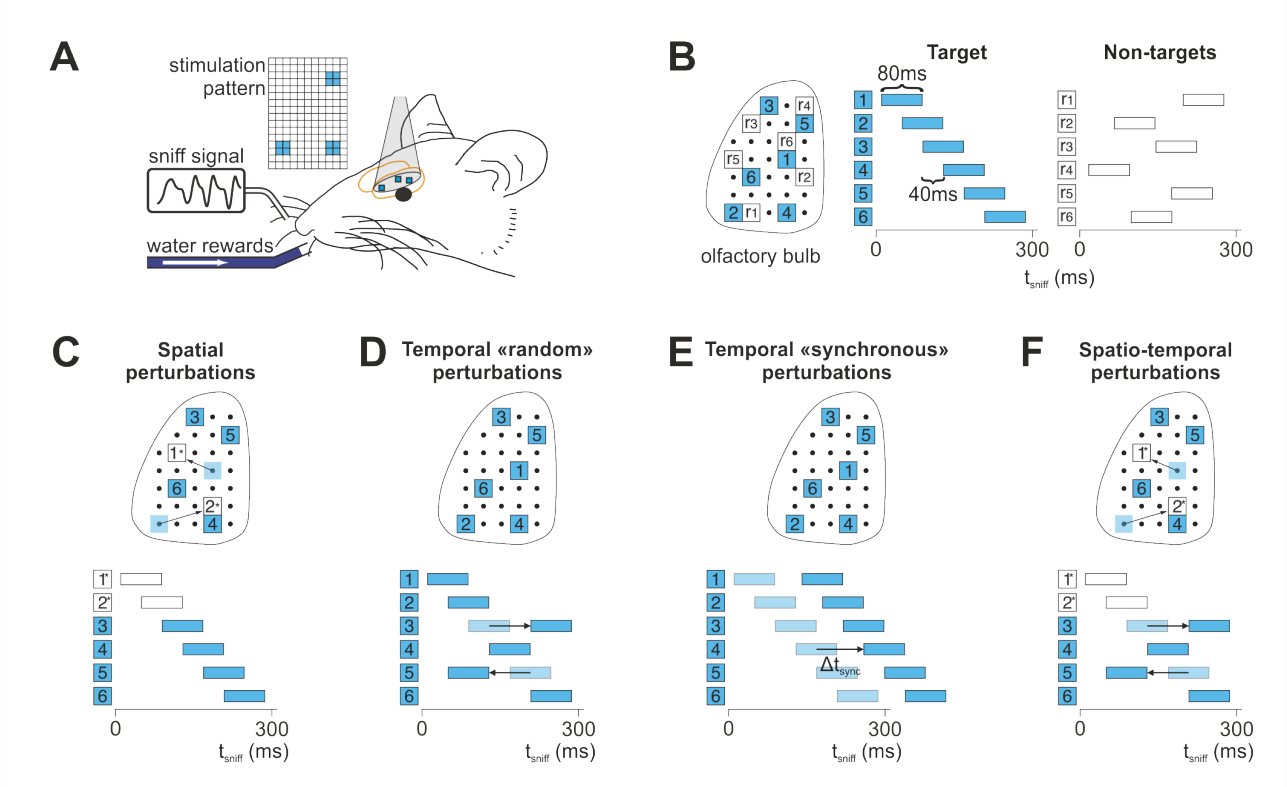

Figure 4.1: **Illustration of the experimental paradigm. A.** Schematic of the experimental setup. Dorsal olfactory bulb (OB) was exposed by a chronically-implanted 3 mm window. Spatio-temporal stimulation patterns, created by a digital micromirror device, were projected onto the OB of a head-fixed mouse in front of lick spouts and a pressure sensor for respiration monitoring. **B.** Schematics for pattern discrimination task. Animals were trained to recognize Target versus Non-target patterns defined on a stimulation grid. Target patterns comprised of six spots activated in an ordered sequence defined in time where 0 marks inhalation onset. Non-target patterns were six off-Target spots, randomly chosen from trial to trial, with randomized timing. **C.** Schematic of spatial perturbations. The identity of one or more Target spots was replaced with off-Target spots. The activation timing was the same of Target pattern. **D.** Schematic of temporal 'random' perturbations. The activation timing of one or more Target spots was randomly shifted in time, while the spots identity was preserved. **E.** Schematic of temporal 'synchronous' perturbations. The activation timing of all the Target spots was shifted in time by the same amount, while the spots identity was preserved. **F.** Schematic of spatio-temporal perturbations. The identity of one or more Target spots was replaced with off-Target spots and the timing of one or more Target spots (the ones that were not replaced) was randomly shifted in time.

randomly selected Non-Target spots, while the activation timing of all the spots was fixed and equal to the Target pattern activation timing (Figure 4.1C);

- *Temporal "random" perturbations*: in these trials the stimulated spots were the same as Target pattern, but the activation onset of one or more spots was changed. The shift in time from the Target activation onset could be positive (towards the end of the sniff cycle) or negative (towards the onset of the sniff cycle) and was randomly chosen in steps of 10 ms. All the activation timing were restrained to the interval from 0 to 300 ms after the inhalation onset ($\sim$ single sniff) (Figure 4.1D);

- *Temporal "synchronous" perturbations*: in these trials the stimulated spots were the same as Target pattern, but the activation onset of all the spots was shifted by the same amount (in steps of 10 ms) (Figure 4.1E);

- *Spatio-temporal perturbations*: in these trials one or more spots were either replaced or shifted in time according to the same rules of the temporal "independent" perturbation trials. No spots were simultaneously replaced and shifted in time (Figure 4.1F).

For any set of perturbations, we measured the fraction of trials where the mouse made a lick choice towards the water spout associated with the Target ('Like-Target' response), as opposed to the Non-target spout. This measurement reflects perceptual distances: the perceived differences between the perturbed and original Target pattern. The larger the perceptual distance, the lower the fraction of choices made to the Target spout.

4.2.2 Model fitting

For the modelling of the data, the full dataset (all sessions from all animals pooled together) was splitted in a training set (75% of the total number of trials) and a test set (25% of the total number of trials). The proportion of each type of trials (Target, Non-Target and each type of Probe trial) was preserved when the full dataset was splitted. To fit each model, we performed 5-fold cross-validation on the training set. The training set, the test set and the

folds for the cross-validation procedure were kept fixed across all the models we considered, in order to avoid that the variability in the results of the fitting could be due to the different trials used to fit the models. We selected as best-fitting parameters those parameters that returned the lowest cross-validated prediction error (measured as percentage of non-correctly predicted trials) averaged across folds. We then evaluated the goodness of fit of each model computing the Brier Score (BS) on the test set. For a binary variable C, the Brier score is defined as

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - c_i)^2 \tag{4.1}$$

where N is the number of samples, p_i is the predicted probability of a positive outcome in the i-th sample and c_i is the observed outcome in the i-th sample. The Brier score ranges between 0 (perfect prediction of the variable outcome) and 1 (totally wrong predictions). To perform statistical comparison of different models, the Brier score was computed on $N = 500$ bootstrap versions obtained from the test set. The bootstrap samples are balanced in the different type of trials and each trial is sampled with repetition from the test set. The Brier score was computed separately for each trial type and for all the trials pooled together.

4.2.3 Logistic regression

We described the data using logistic regression. We set different sets of predictors and compared the goodness of fit of the models in order to find the best fitting model. We used the *glmnet* package in R to fit different logistic regression models. The glmnet package implements elastic net regression, which allows to progressively go from Lasso penalty to ridge penalty by setting a parameter: we empirically set the value of this parameter to 0.5. We compared the goodness of fit (Brier Score computed on the test set and on N=500 bootstrap version of the test set) of all the models versus a null model (only a bias term is used to predict choice) and between them using Anova with correction for multiple comparison (Tukey's Honestly Significant Difference Procedure). Statistical comparisons were performed in Matlab. In the following we report the logistic models we used to fit the data.

Spatial perturbation

To model spatial replacement trials, each activated spot was described by a binary variable, which takes value 1 if the spot has been replaced (in other words it is an off-Target spot), value 0 otherwise. We considered the two following logistic regression models.

(1S) *Linear model.* The first model we considered was

$$logit(p) = \beta_0 + \sum_{i=1}^{N} (\beta_i * x_i) \tag{4.2}$$

Where p is the probability of reporting a Target pattern, x_i is the binary variable corresponding to the i-th spot and β_0 is the bias term and β_i are the weights relative to each variable.

(2S) *Quadratic model.* We also considered a quadratic model, where the effect of replacing multiple spots does not sum up linearly.

$$logit(p) = \beta_0 + \sum_{i=1}^{N} (\beta_i * x_i) + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (\beta_{ij} * x_i * x_j) \tag{4.3}$$

Where the notation is as in model (1S).

Temporal perturbations

Temporal perturbation trials were characterized by changes in the activation timing of Target spots. Given a trial T, we denoted with dt_i the shift in time (with respect to the Target activation time) of the i-th spot of the Target pattern. In the logistic regression models we considered separately shifts towards the end of the sniff cycle and shifts toward the beginning of the sniff cycle, denoted with dt_i^+ and dt_i^- respectively. We considered the two following logistic regression models.

(1T) *Signed time shift.* The first model we considered was

$$logit(p) = \beta_0 + \sum_{i=1}^{N}(\beta_i^+ * |dt_i^+| + \beta_i^- * |dt_i^-|) \tag{4.4}$$

Where p is the probability of reporting a Target pattern, dt_i^+ and dt_i^- are as described above, β_0 is the bias term and β_i^+ and β_i^- are the weights relative to each regressor.

(2T) *Rank order.* We considered a model where the rank order of the spots is used to predict behavior, without any information about the exact timing of activation. Let's denote with R_i, for $i = 1, ..., 6$, the ranking of the Target spots and with Z_i, for $i = 1, ..., 6$ the ranking of the same spots in the considered trial T. In case of ties we consider the average of the Z_i of the spots with the same activation timing. We built a model where the predictor relative to each spot is its change in the rank order, computed as $dr_i = (R_i - Z_i)^2$. The resulting model was:

$$logit(p) = \beta_0 + \sum_{i=1}^{N}(\beta_i * dr_i) \tag{4.5}$$

Where p, β_0 and β_i are defined as in model (1T).

(3T) *Signed time shifts with additive interactions.* In order to check for more complex models containing the relative timing between spots activation, we added to the model (1T) additive interactions terms:

$$logit(p) = \beta_0 + \sum_{i=1}^{N}(\beta_i^+ * |dt_i^+| + \beta_i^- * |dt_i^-|) + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}(\beta_{ij} * |dt_j - dt_i|) \tag{4.6}$$

Where p, β_0, β_i^+, β_i^- and $\beta_i j$ are defined as in model (1T). We considered both the complete model, with all the pairwise interactions, and a reduced model with only the differences in time relative to the first spot.

(4T) *Center of activity with relative shifts.* Another model we considered is the following:

$$logit(p) = \beta_0 + \beta_{SNIFF} * dt_{SNIFF} + \sum_{i=1}^{N}(\beta_i^+ * |\delta t_i^+| + \beta_i^- * |\delta t_i^-|) \tag{4.7}$$

Where dt_{SNIFF} denotes the difference between the global position within the sniff cycle of the considered trial with respect to the global position in the sniff cycle of the Target pattern. We defined the global position of a pattern in the sniff cycle using the temporal *center of activity* of a pattern ($CA = \sum_{i=1}^{N} t_i$, i.e. the temporal average of the activation onset t_i of all its spots). The quantities δt_i^+ and δt_i^- represent the forward and backward difference in the relative timing (discounting for the effect of differences in the global postion in the sniff cycle) of the considered pattern T with respect to the Target pattern $\delta t_i = (t_i^T - CA^T) - (t_i^{target} - CA^{target})$. The quantities p, β_0, β_{SNIFF}, β_i^+ and β_i^- are defined as in model (1T).

(5T) *Center of activity with relative shifts and additive interactions.* As we did for model (1T), we added to model (4T) additive interactions in the form:

$$logit(p) = \beta_0 + \beta_{SNIFF} * dt_{SNIFF} + \sum_{i=1}^{N}(\beta_i^+ * |\delta t_i^+| + \beta_i^- * |\delta t_i^-|) + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}(\beta_{ij} * |\delta t_j - \delta t_i|) \tag{4.8}$$

Where p, β_0, β_{SNIFF}, β_i^+, β_i^- and $\beta_i j$ are defined as in model (1T)

(6T) *Multiplicative interactions.* We considered a last model with quadratic terms. Given the asymmetry in the data, we considered four kinds of quadratic terms: coherent shift of two spots activation timing towards the end of the sniff cycle, coherent shift of two spots activation timing towards the beginning of the sniff cycle, non-coherent shifts that bring two spots closer in time, non-coherent shifts that bring two spots far in time.

$$\begin{aligned} logit(p) = \beta_0 + \sum_{i=1}^{N}(\beta_i^+ * |dt_i^+| + \beta_i^- * |dt_i^-|) + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\big[(\beta_{ij}^{++} * |dt_i^+| * |dt_j^+|) + \\ +(\beta_{ij}^{--} * |dt_i^-| * |dt_j^-|) + (\beta_{ij}^{+-} * |dt_i^+| * |dt_j^-|) + (\beta_{ij}^{-+} * |dt_i^-| * |dt_j^+|)\big] \end{aligned} \tag{4.9}$$

Where p, β_0, β_i^+, β_i^-, β_{ij}^{++}, β_{ij}^{--}, β_{ij}^{+-} and β_{ij}^{-+} are defined as in model (1T)

Spatio-temporal perturbations

Non-target trials and spatio-temporal perturbed trials contained both spatial replacement and temporal shift of the activation onset. The effect of performing both kinds of manipulations could be additive, meaning that we simply need to sum the predictors contained in the best spatial model with the predictors of the best temporal model, or there could be some non-linear effect. For example the effect of replacing a spot and shifting its onset could be less than the sum of the effects of replacing it plus shifting its onset. To test for this, we considered a regression model obtained by combining the spatial and temporal model with a simple sum and another model that accounts for interactions of replacing and shifting the same spot in the same trial. In 'sniff' time coordinates this happens only for non-target trials, but in relative time coordinates this happens also for spatio-temporal perturbation. In case of relative time coordinates, the two models we considered are the following:

(1ST) *Additive model.*

$$logit(p) = \beta_0 + \beta^T_{SNIFF} * dt_{SNIFF} + \sum_{i=1}^{N}(\beta^{T,+}_i * |\delta t^+_i| + \beta^{T,-}_i * |\delta t^-_i| + \beta^S_i * x_i) \quad (4.10)$$

Where the variables notation is the same as in model (1S) and (4T). The weights β^T_{SNIFF}, $\beta^{T,+}_i$ and $\beta^{T,-}_i$ are relative to tempora predictors, while the coefficients β^S_i are relative to spatial regressors.

(2ST) *Interactions model.*

$$\begin{aligned} logit(p) = \beta_0 + [\beta^{ST}_{SNIFF} * dt_{SNIFF} * (1 - \prod_{i=1}^{N}(1 - x_i)) + \beta^T_{SNIFF} * dt_{SNIFF} * \prod_{i=1}^{N}(1 - x_i)] + \\ + \sum_{i=1}^{N}[(\beta^{T,+}_i * |\delta t^+_i| + \beta^{T,-}_i |\delta t^-_i|) * (1 - x_i) + (\beta^{ST,+}_i * |\delta t^+_i| + \beta^{ST,-}_i |\delta t^-_i|) * x_i + \beta^S_i * x_i] \end{aligned} \quad (4.11)$$

Where the notation is the same as in model (1S) and (4T). The interaction model allows to have a non-additive effect of simultaneously replacing and shifting the same spot. It

does not account for spatio-temporal interactions deriving from manipulations of different spots.

4.2.4 Perceptual metric

We built a model that maps OB activity patterns to choice through the definition of a perceptual metric on the OB activity space. Given any two OB activity patterns, the model computes the perceptual distance between them and passes this quantity through a logistic function in order to estimate the probability of reporting a match between the two patterns. We started from the well formulated van Rossum distance for spike trains in order to build the metric that maps neural activity patterns to olfactory perceptual representation. In order to fully describe animal perception, the metric needs to account for differences in activation timing and spot identity. Furthermore, the representation of activation timing needs to contain both the position of the pattern within the sniff cycle and the exact of single spots within the pattern. On the top of this, the differences in patterns spatial and temporal features must be modulated by primacy. The mathematical function describing a metric that satisfies such requirements is described in Appendix A. We describe here the main features of this metric. A neural pattern T is represented as a set of signals coming from N different channels, where N is the total number of spots (or glomeruli). When a spot is activated, it generates a transient response in the corresponding channel. Such response is modelled with an exponential decaying kernel with exponential time constant τ_{act}. The amplitude of the signal is modulated by an exponential decaying function with exponential time constant $\tau_{primacy}$ and onset corresponding to the earliest activation timing of the pattern spots. The amplitude modulation of the single channel responses models the primacy effect. Given two neural patterns T and S, their distance is given by the sum of two components. The first component is a function of the difference of their position in the sniff cycle. We tested different ways of representing the position of the patterns in the sniff cycle:

- an average of all the activation onsets;

- a primacy-weighted average of the activation onsets (center of activity of the function given by the summation of the waveform of all channels);
- the earliest activation timing of the pattern;
- the center of activity of the cross-correlogram function (obtained computing cross-correlation for single channels and then summing up the cross-correlograms before computing the center of mass).

If not specified, we used the primacy-weighted average of the activation onsets to define the pattern position in the sniff cycle. This component is passed through a saturation function whose temporal course is modulated by τ_{glob} and whose upper bound is given by λ_{glob}. The second component of the distance represents the total difference in the activation timing of the different channels, after discounting for the effect of changes in the global position in the sniff cycle. This quantity is computed by first aligning the pattern representations to their center of activity and then computing, for each channel, the difference in the area covered by the two pattern waveforms (within-channel difference). The sum of the within-channel differences is the second component of the perceptual distance. A further extension of the metric definition allows to account for the role of spots identity in perception. By introducing a parameter θ, that regulates the contamination across channels, it is possible to switch from a summed population code to a labelled line code, as in [29].

Overall, the metric depends on five parameters:

- $\tau_{primacy}$: exponential time constant of the primacy curve. Small values of this parameter result in a large primacy effect, where the effect on perception of changing later spots is negligible. Large values of this parameter tend to reduce the primacy effect, which is absent when the primacy curve becomes flat.
- τ_{act}: exponential time constant of the activation kernel. Large values of this parameter result in a gradual effect of temporal shift, with perceptual distance slowly increasing for increasing shifts. Small values of this parameter result in an abrupt effect of temporal

shifts, with perceptual distance suddenly increasing for small shifts and then remaining constant from further deviations from a Target.

- τ_{glob}: time course of the saturation of the effect of global shift. Small values of this parameter result in an early saturation of the effect.

- λ_{glob}: upper bound of the global distance component. This parameter modulates the trade-off between the two components of the perceptual distance (difference in the center of activity and within-channel differences).

- θ: this parameter models the relevance of spot identity in the code. It ranges from 0 (all the channels are collapsed together before comparing activation timing) to $\frac{\pi}{2}$ (separate channels activation timing are compared independently and then summed up).

We show in Appendix A that all the metric properties are preserved.

To fit the perceptual metric we used the same training and test sets as for logistic regression. We defined a grid in the 5-dimensional parameter space and, for each grid point, we computed perceptual distances on the training set trials. We than mapped perceptual distances to behavioral choice using logistic regression and a 5-fold cross-validation procedure. The parameters for the logistic regression were obtained using the *glmnet* package in R (elastic net penalty parameter = 0.5). We considered as best fitting parameters for the perceptual distance those parameters in the grid that returned the lowest prediction error averaged across folds. For these parameters we ran again logistic regression on the total training set to estimate the logistic regression weights. We evaluated the goodness of fit of the perceptual metric model and the comparison across different models using the test set in the same way we did for logistic regression. We want to underline here that the grid-search procedure is a brute-force approach and does not guarantee the convergence to the optimal solution. However, the prediction accuracy we got with the grid-search procedure is very close to the one of the logistic regression fit, which uses optimized algorithms to estimate the best fitting coefficients. This made us confident that our approach lead us close to an optimal solution.

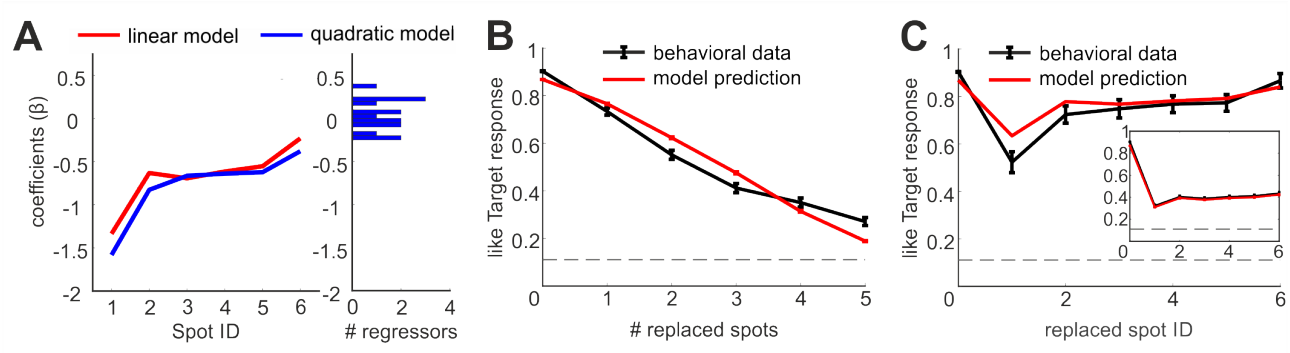

Figure 4.2: **Logistic regression accuracy in behavioral outcomes prediction for spatial perturbations trials. A.** Coefficients of logistic regression models. Left. Coefficients of linear terms for both the linear (red) and the quadratic model (blue). The larger magnitude of coefficients related to early activated spots suggests a primacy effect in behavioral readout. Right. Distribution of the coefficients of quadratic terms. The concentration of coefficients around zero suggests that the behavioral readout of spatial features is linear. **B.** Effect on animal behavior of the replacement of a different number of spots. In (**B**) and (**C**) behavioral data (black) are plotted together with the predictions of the linear regression model (red). The dashed gray line shows the behavioral performance for Non-target trials. **C.** Effect on animal behavior of the replaced spot identity for single spot replacement trials (Inset. Same for multiple spots replacement trials).

4.3 Results

4.3.1 Spatial coding

We first analysed trials with spatial perturbations in order to investigate the role of spot identity in odor coding. From behavioral plots (Figure 4.2C), we observed that the replacement of earlier activated spots has a stronger effect on behavior than the replacement of later activated spots, resulting in a higher probability of reporting a match with the Target pattern of trials where latest spots were replaced. When replacing multiple spots, furthermore, the discriminability of patterns from the Target decreases almost linearly as the number of replaced spots increases, independently from spot identity (Figure 4.2B). The observed decrease in discriminability can be explained either by integration of the signal across spots or by an increased probability of replacing earlier activated spots as the number of replaced spots increased. We ruled out the second possibility by observing that replacing more spots in the same pattern, after having conditioned on the replacement of a given spot, increases the perceptual discriminability from the Target (inset in Figure 4.2C). This suggests that the animal is integrating information in a

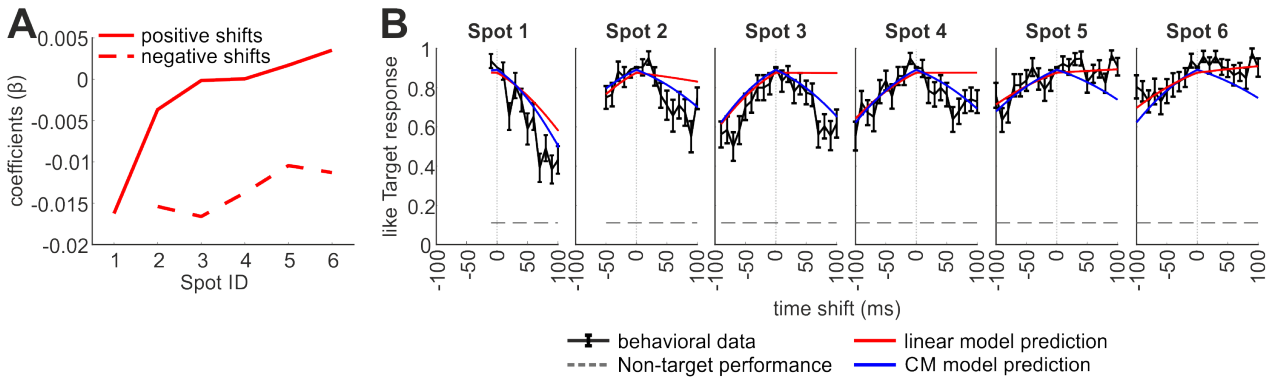

Figure 4.3: **Logistic regression accuracy in behavioral outcomes prediction for temporal "random" trials. A.** Coefficients of linear logistic regression model for shifts towards the end (continuous line) or the beginning (dashed line) of the sniff cycle. The larger magnitude of coefficients related to early activated spots and, in general, of coefficients related to shifts towards the beginning of the sniff cycle suggests a primacy effect in behavioral readout. **B.** Effect on animal behavior of the shift of each spot (only trials with single spot shift). Behavioral data (black) are plotted together with the predictions of the linear regression model (red) and of the center-of-activity regression model (blue). The dashed gray line shows the behavioral performance for Non-target trials.

window of time comprising more than a single spot and that information about the first spots is not sufficient to describe behavior. Such qualitative observation were quantitatively described through logistic regression: a linear model is as good in explaining behaviour as a quadratic model (Brier score linear = 0.1574±0.0002, Brier score quadratic = 0.1597±0.0003, p-value = 9.56e-10, detailed results in Supplementary Figure B.1), the coefficients of the quadratic terms have small magnitudes and are centered around 0. Furthermore, the magnitude of the coefficients is larger for earlier activated spots (Figure 4.2A), in agreement with their stronger effect on behavior. We can therefore conclude that the readout of spatial features is linear and weighted by primacy.

4.3.2 Temporal coding

We then analysed the readout of temporal features of neural activity (Figure 4.3). Behavioral plots of single spot shift trials revealed two main features: i) changing the activation timing of earlier spots has a stronger impact on perception, resulting in a lower discriminability threshold; and ii) the effect of moving the activation onset toward the onset or the end of the sniff cycle

is not symmetric. We quantified such observations with a simple logistic regression model, that contained as predictors the shifts in time from the Target pattern, considered separately according to their direction (toward the beginning or the end of the sniff cycle). In agreement with qualitative observations, we obtained coefficients with larger magnitude for perturbations of earlier spots and for shifts toward the first part of the sniff cycle (Figure 4.3A). From this we concluded that, as for spatial features, the relevance of temporal features is weighted by primacy.

However, the model we used fails in explaining the results of global shift trials (Figure 4.4A). The main difference between the two types of perturbations is that in the second type the internal temporal structure of the pattern is preserved, in the sense that the relative timing between each pair of spots is constant. In order to check whether this internal structure could be a candidate to play a role in perception, we splitted the perturbed trials in trials where the ordering of spots was preserved and trials where the order of spots was changed. This is a rough approximation of trials with small and large changes in the internal temporal structure. As we had hypothesized, we observed that the discriminability from the Target pattern was larger when the order of spots was changed (Figure 4.4B), suggesting that changes in the relative position within the sniff cycle are read out and are sufficient to drive behavior.

Therefore, we formulated different logistic regression models in order to explain all the observed data (details of the model fits in Appendix B). We considered a rank model, that discarded any information about activation timing considering only changes in the activation order of spots. Such model failed in explaining the global shift trials and was less accurate than the linear timing model in predicting the behavior for temporal 'random' perturbations (Brier score linear = 0.2175±0.0005, Brier score rank order = 0.2346±0.0006, p-value = 6e-08, detailed results in Supplementary Figure B.2 and Supplementary Figure B.3). This provides evidence for a read out that takes into account both absolute timing in the sniff cycle and relative timing between spots. We therefore tried to fit a logistic regression model that included relative shifts in the form of relative differences between pairs of spots. Such model was able to improve the description of behavioral data for 'synchronous' shift trials (Brier score linear = 0,2079±0.0004, Brier score difference = 0.1990±0.0004, p-value = 6e-08, detailed

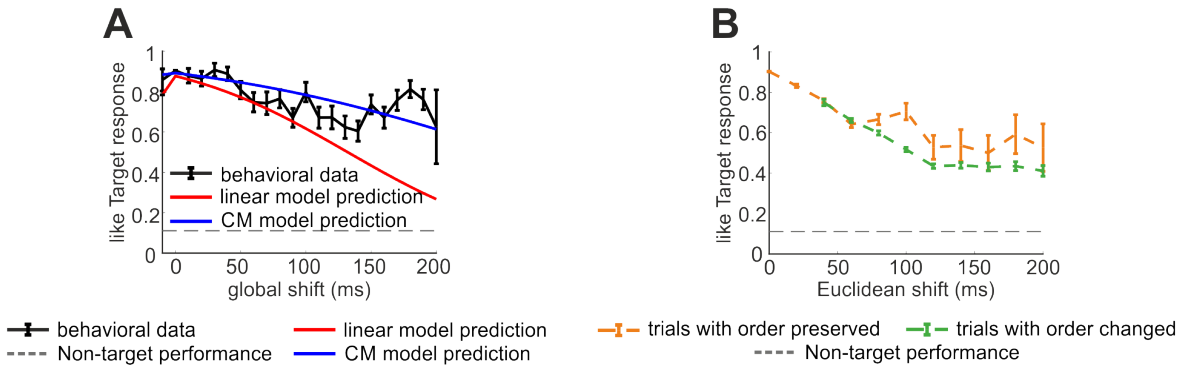

Figure 4.4: **Logistic regression accuracy in behavioral outcomes prediction for temporal synchronous trials. A.** Effect on animal behavior of the simultaneous shift of all spots. Behavioral data (black) are plotted together with the predictions of the linear regression model (red) and of the center-of-activity regression model (blue). The dashed gray line shows the behavioral performance for Non-target trials. **B.** Effect on animal behavior of multiple spots shift as a function of the total euclidean shift from Target pattern timing for trials where the order of the spots is preserved (orange) or not preserved (green). The lower probability of Target detection for trials where the order is not preserved suggests a role of the internal pattern structure for behavioral readout

results in Supplementary Figure B.2 and Figure B.3), but not for temporal 'random' trials (Brier score linear = 0.2174±0.0005, Brier score difference = 0.2179±0.0005, p-value = 0.99, detailed results in Supplementary Figure B.2 and Figure B.3). We then formulated a simpler model, including one term that represents the position of the pattern in the sniff cycle and other terms representing the internal temporal structure of the pattern. Such a model was better in describing the data (Brier score linear = 0.1720±0.0003, Brier score difference = 0.1576±0.0003, p-value = 6e-08, blue lines in Figure 4.3 and Figure 4.4), suggesting that the readout of activation patterns uses two separate features: the first is the broad position of neural activation locked to the sniff cycle and the second is the precise internal temporal structure of neural activity. From this, we concluded that temporal features of neural activity are read out in two different temporal reference frames, a sniff-locked one and an internal one.

4.3.3 Perceptual metric: a unifying model

After having characterized the readout using logistic regression, we tried to build a metric that could lead to such readout (Figure 4.5). We considered the spots on the OB as distinct channels.

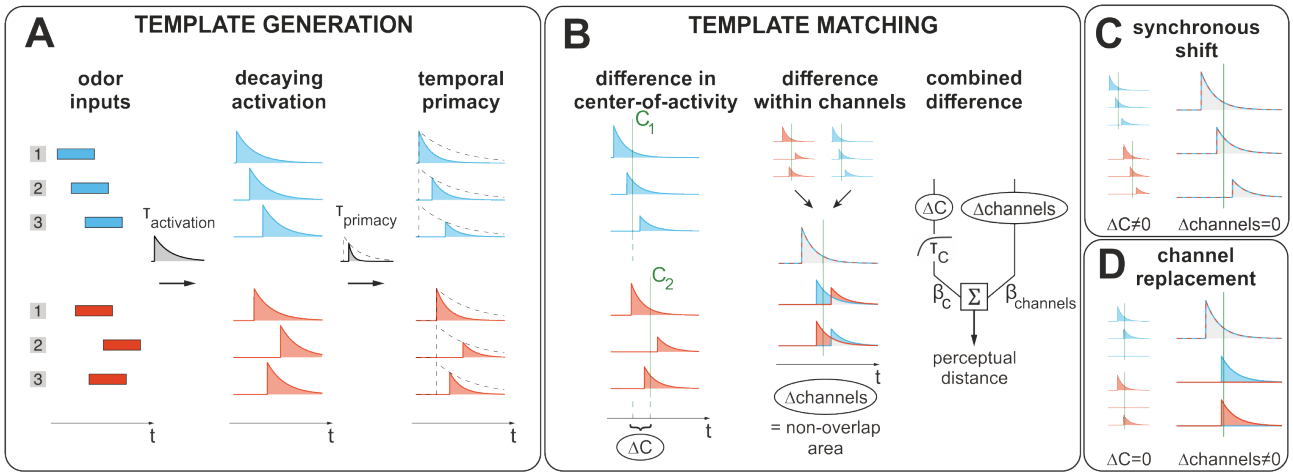

Figure 4.5: **Illustration of perceptual metric model for OB readout. A.** Each odor input is defined by the activation of different channels with a precise timing (blue and red squares). In order to evaluate the difference between patterns, each pattern is transformed into its waveform representation through convolution of each channel with an exponential decaying kernel. The amplitude of the kernel is modulated by an exponential decaying function starting at the pattern onset. The decay time of the kernel and of the primacy curve are two model parameters. **B.** For each pattern, its center of activity is computed. The center of activity represents the broad position of the pattern in the sniff cycle. Each pattern is aligned to its center of activity and the non-overlapping area between the "centered" patterns is computed channel-by-channel and then summed up. The distance between the two patterns is a weighted sum of the total area under the waveforms and the absolute difference between the centers of activity, filtered through an exponential function in order to model a saturation effect. **C.** Example of the metric computation for patterns where all the spots are synchronously shifted in time. In this case the difference in the center of activity is larger than zero, while the channel-by-channel difference is zero. **D.** Example of metric computation for trials with spot replacement. In this case the difference in the center of activity is zero, while the channel-by-channel difference is larger than zero.

The activation of a channel generates a transient response, represented by an exponential decaying kernel, whose exponential time constant is fixed across channels and across time and whose strength is modulated by primacy. Each pattern is therefore characterized by the waveforms it generates in each channel. From such a representation, animals extract two quantities: a global positioning of the pattern in the sniff cycle and a centered representation of the pattern, meaning that such representation discounts for the global positioning in the sniff cycle. A measure for the discriminability between two patterns is given from a weighted combination of the difference in global position in the sniff cycle and the difference in the centered pattern representation integrated across all channels. Fitting this model to the data, we observed that, after the centering step, there is no contamination across channels (represented by the parameter $\theta = \frac{\pi}{2}$), suggesting an independent readout of each channel. The exponential time constant of the activation kernel was ca. 70 ms, which is in line with the activation time of M/T cells, obtained from electrophysiological recordings [4]. The primacy kernel is not flat, suggesting that primacy coding is effectively implemented by the brain. The exponential time constant of the primacy kernel is around 200 ms. We showed in Figure 4.5 that this model is capable of predicting animal behavior with a performance comparable to the best logistic regression model we considered (Brier score metric = 0.1592±0.0002, Brier score logistic regression = 0.1601±0.0002, p-value = 7e-87) (Figure 4.6), with a lower number of parameters: for the perpetual metric we had to fit n=7 parameters (five parameters for the metric and two for the logistic mapping to behavior), while for the best spatio-temporal logistic regression model we had to fit n=20 parameters (the bias term, the weight of the center of activity predictor, six coefficients for the spatial predictors and twelve coefficients for the signed temporal coefficients). Furthermore, the number of parameters of the metric is fixed and independent from the number of spots, while the number of regressors of the logistic regression model (and therefore the complexity of the model) is a function of the number of spots ($n_{params} = 2 + 3 * n_{spots}$).

We further tested the model we built by progressively modifying or removing processing components. We first compared different techniques for the computation of the positioning in the sniff cycle and the centering of the patterns (detailed results in Supplementary Figure B.4, Supplementary Figure B.5 and Supplementary Figure B.6). The best fitting techniques used

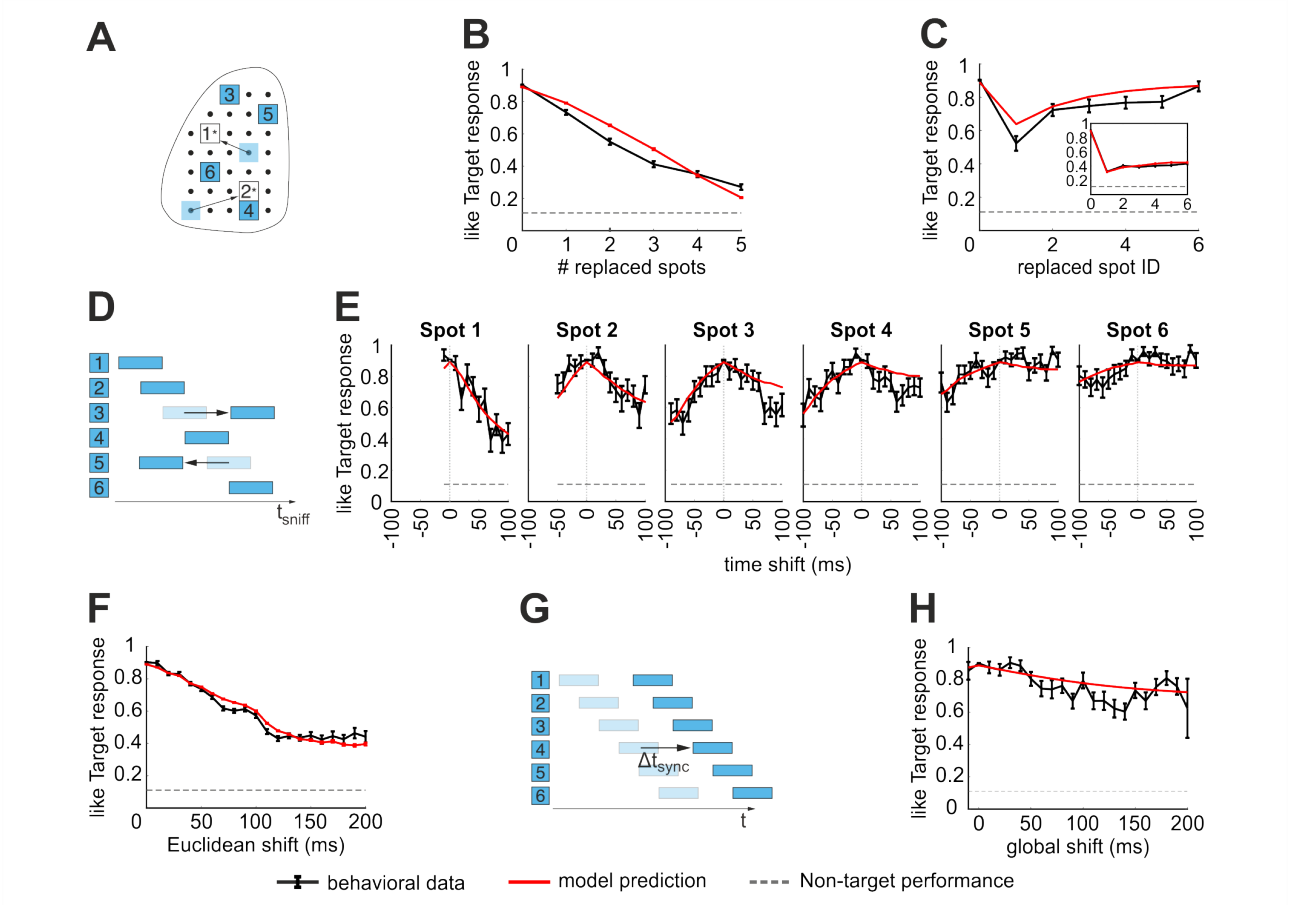

Figure 4.6: **Perceptual metric accuracy in behavioral outcomes prediction. A.** Illustration of spatial replacement trials. **B.** Effect on animal behavior of the replacement of a different number of spots. In (**B**), (**C**), (**E**), (**F**) and (**H**) behavioral data (black) are plotted together with the predictions of the metric model (red). The dashed gray line shows the behavioral performance for Non-target trials. **C.** Effect on animal behavior of the replaced spot identity for single spot replacement trials (Inset. Same for multiple spots replacement trials). **D.** Illustration of temporal 'random' trials. **E.** Effect on animal behavior of the shift of each spot (only trials with single spot shift). **F.** Effect on animal behavior of multiple spots shift as a function of the total euclidean shift from Target pattern timing. **G.** Illustration of temporal 'synchronous' trials. **H.** Effect on animal behavior of the simultaneous shift of all spots.

the activity pooled across channels to define a *center of activity* of the neural pattern. The center of activity was then used to discount the position in the sniff cycle from the temporal information. Alternative techniques, that take into account only the earliest activation timing or the non-weighted activity of the channels in order to implement a centering procedure, led to a loss in the predictive power of the model. We also tried to remove the centering procedure or to perform it without considering the difference in the global activation timing as a component of the perceptual distance. Both methods impaired the model performance. This suggests that the centering procedure plays a key role in perceptual discrimination. We therefore selected the weighted *center of activity* technique for patterns centering and removed or modified single components of the metric definition (detailed results in Supplementary Figure B.7, Supplementary Figure B.8 and Supplementary Figure B.9). First, we changed the activation kernel by imposing a fast decay and a slow decay (exponential time constant = 10 ms and 200 ms, respectively). Then, we removed the primacy effect by considering a flat primacy curve. Finally, we investigated the effect of spots identity on neural coding by setting the parameter θ to 0 (summed population code) or by fitting it as an additional parameter of the model. We observed that modifying any of the model components led to a statistically significant loss in model accuracy in predicting behavior. The largest loss was caused by setting the parameter θ to 0. This provides further evidence that the code used for perceptual readout is not a summed population code, but uses the identity of the activated glomeruli. Also shortening the decay of the activation kernel led to a strong impairment in prediction accuracy: in details, effects of spatial perturbations were underestimated, while the predicted sensitivity to temporal perturbations was larger than the observed one.

4.4 Discussion

In this work we introduced a novel approach to study the neural code: the combination of an innovative intervential technique, such as optogenetics, and mathematical tools allowed us to investigate the role of spatial and temporal neural features in perceptual representation in a systematic and independent manner. Optogenetic intervention in awake trained animals allowed

us to build a large dataset with controlled perturbations of neural features and behavioral data. In order to unveil the mechanisms of olfactory code in the mouse olfactory bulb we built a computational model that maps neural activity patterns into a perceptual space. The distance between two patterns in the perceptual space can be used to predict the animal behavior and can be considered as a proxy of the animal internal representation of the stimulus. In the definition of the perceptual space we tried to use biologically plausible computations, but we do not claim that they are carried out by the brain, neither we try to find the biological mechanisms underlying such computations. On the analytical side, we showed that the function we introduced to map neural patterns to perception satisfies the mathematical conditions that define metrics and, at least theoretically, allows comparing any neural activity pattern. On the experimental side, modelling perception with this approach resulted in a prediction accuracy similar to a classical method such as logistic regression, but provided the advantage of being characterized through a smaller number of parameters that is independent from the number of considered channels (spots or glomeruli). This makes our method less prone to overfitting problems in the parameters estimate. Furthermore, the mechanisms used to define the metric are more interpretable than the logistic regression equations and provide some clues about possible processing and readout mechanisms.

We grounded our work on the neural metrics defined for the study on neural encoding in [74, 75, 53] and extended in [3, 29]. As in those works, we represented neural activity using exponential decaying kernels, that could be considered a first approximation of post-synaptic responses and, more important, lead to changes in the neural activity representation only in the time window following a stimulus, thus preserving causality.

The STM model resolves several open issues in olfactory coding. Confirming results already present in literature, our approach provided evidence for a primacy effect in neural code where earlier-activated glomeruli have larger effects on perceptual responses. Primacy has been suggested as a strategy for animals to recognize the same odor across varying concentrations, as early-activated glomeruli remain stable across different concentrations. Animals trained to recognize odors across varying concentrations were impaired during coarse optogenetic disruption, but only when the disruption occurred in the initial 100ms of inhalation [77].

In the present study, animals were trained to recognize a single Target synthetic odor, and displayed a primacy effect without explicit training to "concentration-variants" of the Target. Hence, the preferential weighting of early inputs occurs under different task demands, suggesting that primacy is a fundamental property of the olfactory system. We found evidence for a readout of both spatial and temporal features of neural activity. The readout of spatial features can be modelled as a linear summation of independent channels. It has been proposed that odors are encoded in unique spatial combinations of glomeruli [40], but there are multiple possible combinatorial coding schemes which have not been tested. For example, one extreme possibility is a barcode representation, where any slight change to the combination leads to a completely un-related odor, maximizing the systems representational capacity for odors. Instead, we found a perceptual readout of spatial patterns that is linear. This is consistent with other studies in the literature [41, 66, 23]. A linear readout implies that two patterns will generate odors that are perceptually similar depending on the degree of glomerular overlap. While a linear readout has lower representational capacity, it may explain the generalization of odor percepts across varying concentrations or backgrounds, where activated glomeruli differ slightly.

Our work causally establishes the role of temporal sequences in odor perception, which has long been hypothesized [11, 63] but not directly tested: temporal codes are not theoretically required to support odor perception [37]. We presently demonstrated that mice trained to recognize a synthetic odor pattern, use temporal sequences in odor recognition even though other (spatial) cues are sufficient to solve the task. Surprisingly, we further found that these sequences were defined with relative latencies within a pattern [35], as opposed to latencies with respect to sniff as previously proposed [65, 11, 63, 62]. Pattern-referenced timing may reflect how inputs compete or are integrated downstream, as competition or integration can depend on temporal proximity between inputs [24, 18, 2]. We found a weak perceptual effect for changing the overall position (center-of-activity) of the pattern within sniff, possibly arising from weak modulation of glomerular activity by sniff-coupled mechanosensory responses [32]. Alternatively, mice may be using the overall position of the pattern in sniff as a weak cue to solve the task, as the average randomly-generated Non-target pattern has a different position within sniff compared

to the Target pattern. Leveraging the flexibility allowed in the metric definition, we tried to deepen the understanding of how the system extracts the information about the position of neural activity in the sniff cycle. A possibility is that the earliest activated channel generates a feedback inhibition process and mice use this mechanism to "center" neural activity. This would require knowledge only of the earliest activation time to reference a pattern to the sniff cycle. A second hypothesis is that mice integrate the signals of multiple channels to compute the position of neural activity relative to sniff. We found that a model that integrates information across channels is more successful in predicting behavioral performance, providing evidence for the latest hypothesis. However, how such infromation is integrated across channels remains an open question: the information might be pooled after the pairwise comparison of single channels (like in a template-matching approach based on cross-correlations) or alternatively the activity of all the channels might be pooled together before the computation of the positioning in the sniff cycle. Disambiguating between such hypotheses requires further investigations of the system.

We cannot rule out the possibility that fibers of passage were activated during our spot stimulation. However, this effect is likely to be weakMT cells were activated by few localized spots at the laser intensity chosen for our experiments. Another study also using OMP-ChR2 mice, with similar stimulation size, duration and intensity have also found negligible effects of fibers of passage[24]. Increasing the stimulation intensity led to more apparent recruitment of fibers of passage, reflected in increased number of active spots located anteriorly (unpublished). Furthermore, we observed negligible non-linear interactions between spots during spatial replacements, which is not expected if there is widespread co-activation of glomeruli from fibres of passage.

Another limitation of the study is that, in order to test our mathematical model, we have used synthetic stimuli, that may not capture the full complexity of glomerular activity evoked by natural or naturalistic odors. Our claim is not that these synthetic stimuli are direct proxies for naturalistic stimuli, rather they afford well-controlled experiments with precise parameterization, that can be used to test basic principles of the neural code. We instead view the synthetic approach as complementary to approaches using more naturalistic stimuli,

that are necessary for the validation of the hypotheses formulated about readout mechanisms. In other well-studied sensory systems, foundational understanding of sensory processing has been built upon synthetic, reduced stimuli [31], while naturalistic stimuli have been used to test and refine foundational models [55]. In principle, the metric we constructed would be a valuable tool to predict the discriminability between different odorants, but such a statement needs to be tested through experiments.

Future directions of the project aim at describing the neural response to natural stimuli, for example by building up a dictionary of OB responses to different odorants, and then mapping these responses to behavior using the perceptual metric we presented. One possibility to achieve this goal consists in testing the accuracy of our method in an indirect way. First, we run a series of experiments where we build a map from a set of natural stimuli to neural activity patterns. Then, we use the perceptual metric to predict behavior and we check whether the predicted discriminability is in accordance with the behavioral outcomes we get in a separate series of experiments, where we measure the behavioral response of animals to the set of same stimuli. A second chance, more challenging from the experimental point of view, would be to build an experiment where we simultaneously record neural responses and the behavioral reports. This would allow to test the accuracy of the model on data with a trial-to-trial correspondence between neural response and behevioral response, which would theoretically allow to uncover also the role of noise correlations in the neural code. In case the model is effective in describing the animal internal representation of perceptual similarity, than we could use such a tool to build future experiments, where we can a priori control and manipulate the difficulty of the discrimination task.

Chapter 5

Conclusion

5.1 Summary of the work

In this work we introduced a novel approach for the investigation of behavioral readout, based on the mathematical concept of metric. Given a set of elements (for example a set of stimuli or of neural responses), a metric defines a relationship between each pair of elements, allowing for a rigorous comparison and classification. The only constraints for the definition of a metric are few mathematical conditions that must be satisfied. For this reason, the metric approach is very flexible and, in principle, allows to capture also complex relationships.

The application of metrics to neuroscience is not a novelty. Thanks to the definition of spike train metrics, interesting insights have been gained in the understanding of stimulus encoding ([74, 3]). However, the same approach has been rarely applied to the investigation of behavioral readout. To our knowledge, the only example of application of such approach can be found in [71]. For this reason, we wanted to formally define the application of metric spaces to the study of behavioral readout and show an example of application.

Our method consists in the embedding of neural activity patterns into a metric space, through the definition of a distance function that assigns to each pair of neural activity patterns a positive quantity. We called this quantity *perceptual distance*, because it should reflect how

different neural patterns are perceived by subjects. The approach of defining metrics to study behavioral readout, besides being formally well defined, also presents the important advantage of being general and flexible (it depends on the way the metric is defined) and therefore could be applied to the study of different systems without any a priori limitation. Another important feature of this method is that it is based on single trial measurements. This allows to model also trial-to-trial variability and to take into account the role of noise correlations in the neural code.

We showed an application of the presented approach to the study of perceptual readout of OB activity in mice. Combining optogenetic intervention with this new mathematical method, we were able to investigate the causal role of spatial and temporal features of OB activity on behavior. We modelled the effect of perturbing such features and showed that both spatial and temporal features are read out by the system to guide behavior and that neural activity readout is modulated by primacy, which means that early activity has a stronger impact on behavioral outcome. We also found that spatial and temporal features are processed differently by the neural system. For spatial features a linear model is capable of predicting animals behavior, while a more complex model was necessary to explain the readout of temporal features. Temporal patterns of neural activity are compared using two reference systems: the first reference is realtive to the sniff cycle and we obtained it by computing, for a given neural pattern, an overall center of the activity and extracting its position with respect to the sniff cycle onset. The second reference system is internal to the pattern and reflects the activation timing of each channel (spot or glomerulus) relative to the others. We found evidence that mismatch between patterns in both reference systems were necessary to predict correctly behavioral outcomes, with a stronger weighting for differences in the patterns internal structure. We tested the robustness of the results obtained using the perceptual metric approach by applying logistic regression to the same data. In order to achieve a prediction accuracy similar to perceptual metric, we needed a logistic regression model with a number of parameter at least three times larger than the metric model we introduced. Therefore, we concluded that perceptual metric definition can be considered as a valid alternative to classical approaches for the study of behavioral readout, presenting the advantages of a lower complexity (in terms of

number of parameters) and a better interpretability (paper in preparation, abstract in Appendix C). Future developments of this work consist in investigating the temporal sensitivity of the readout. It has been show that different brain areas are characterized by different temporal sensitivity, from few milliseconds to tens of milliseconds [79, 49]. In this work we have not investigated the temporal precision of the readout in a systematic manner, but the causal approach and the choice of perturbation patterns would be very suitable for this kind of study, that we want to carry out in the future.

5.2 Link to other works and future directions

In this work we have introduced a new approach to the study of perceptual readout. In order to validate the presented approach, we applied it to model mice responses to artificial optogenetics stimulation of the OB. The synthetic stimuli we used did not resemble OB activity patterns evoked by natural or naturalistic odors, but were designed in order to explore in a systematic way the space of neural patterns and to question independently the causal role of spatial and temporal features in behavioral readout. This approach is very powerful in investigating the sensitivity of the system to neural features, but cannot be used to infer conclusions about the readout of neural activity generated by natural stimuli. For this reason, a complementary approach for an ethological validation of the perceptual metric approach requires the application of such method to experiments with natural stimuli. We propose here two possible experimental designs that complement the work presented in Chapter 4. A first approach consists in: i) recording OB activity in response to odors without tracking behavior, and separately ii) recording the behavioral outcome of an odor discrimination task with the same odors as in i) without monitoring the neural activity. This approach allows for an indirect validation of the perceptual metric model and does not leverage the full potentialities of this method, because it does not provide a way to model trial-to-trial variability neither to conclude anything about causal relationships. An alternative approach, the ideal one, consists in recording OB activity in response to real odors, while animals are performing an odor discrimination task. In this way we would be able to model, in a trial-to-trial fashion, causal relationships between stimuli, neural

activity and behavioral outcomes. Introducing perturbations of the stimulus-evoked neural activity (in other terms "biasing" sensation), we might draw conclusions about which neural features are causally relevant to perception. Our collaborators in New York are equipped with experimental setups that allow to continue the investigation of perception in both the suggested directions. We are already collecting data to study the encoding of naturalistic odors in the olfactory bulb (with high spatial and temporal resolution) and we are able to manipulate neural activity in awake animals with single neuron resolution [38]. We are currently developing an experimental approach that could allow to bias sensation in a reasoned way, by activating or silencing putative perceptually relevant neurons.

Another important aspect to consider about the experimental paradigm we used to validate our model is the task performed by the animal [34]. We modeled animal behavior in a 2AFC task and we chose to use logistic regression to link the metric measurements to behavioral outcomes. In future, to assign an ecological validity to the proposed approach, we would need to apply the metric model to other behavioral paradigms. Indeed, in a more naturalistic environment, animals might face multiple alternative choices (instead of binary ones). A few mathematical approaches already exist to predict multiple choice, like multinomial logistic regression, nested logistic regression, multinomial probit regression or other multiclass classification methods. Such methods can be easily combined with the metric definition and, depending on the kind of task, one or more of these approaches might be implemented.

The suggested experiments, with natural stimuli and naturalistic-like behavioral outcomes, a part providing a way to validate the metric approach, are suitable to thoroughly investigate the perceptual process from stimulus to behavior. As pointed out many times, perceptual readout is only a single step of this complete process and its understanding is not sufficient to unveil the full mechanisms underlying perception. As Panzeri et al. highlighted in [50] what really matters to understand neural codes is the information about stimuli that is encoded in neural activity and simultaneously read out to drive behaviour at single trial level. Only experiments with simultaneous behavioral report and neural recordings allow to study this link on a trial-to-trial basis.

One of the main challenges to conduct experiments with simultaneous neural recordings and behavior is the necessity of recording neural activity in awake subjects. Developments and advancements in neural recording techniques are pointing to this direction, providing tools to record neural data with always higher quality. Among all the develeped recording techniques, in the last years two-photon calcium imaging has become one of the most promising tools to record neural activity [20]. Two-photon calcium imaging records calcium activity of neurons expressing calcium indicators, that are molecules that change their fluorescence in response to calcium binding. A class of calcium indicators are the GECIs (genetically encoded calcium indicators), that are expressed only from specific types of cells (selected by researchers) and present the advantage of being suitable for long-term studies. Two-photon calcium imaging measures the changes in the fluorescence of neurons expressing GECIs and uses such measurements as a proxy of the spiking activity. It has been shown [12] that the latest developed families of GECIs provide single spike resolution in the calcium signal deconvolution (that is the inverse process of reconstructing the spiking activity that generates a given calcium signal). One of the main strengths of two-photon calcium imaging is the possibility of recording up to hundreds or even thousands of neurons with single-cell resolution. This makes two-photon calcium imaging an appealing tools for neural population studies. The drawback of such a high spatial resolution is the limited temporal resolution of classical two-photon calcium imaging. Furthermore, as most imaging techniques, calcium imaging is limited to superficial brain areas because of the scattering caused by brain tissues. In collaboration with Andrea Sattin and Marco Brondi from the Optical Approaches to Brain Function Laboratory at IIT in Genova, under the joint supervision of Tommaso Fellin, we worked on the development of technicques to improve the quality of two-photon calcium imaging recordings, in order to overcome the presented limitations and fully exploit the potentialities of this technique.

The first strong limit of two-photon calcium imaging, the low temporal resolution, is due to the way the signal is acquired. Classical two-photon calcium imaging collect signals scanning the entire field of view (FOV) pixel-by-pixel in a raster mode. This approach is highly time-consuming and implies always lower temporal resolution with increasing FOV size (and consequently of the number of imaged neurons). Furthermore, with such a low temporal

resolution, the single spike accuracy potentially provided by the latest developed GECIs could never be obtained. In order to overcome this limit, a number of techniques have been proposed (see [59] for a review), for example the use of resonant galvos, of acousto-optic deflectors (AODs) [19], parallel scanning methods [51], holographic excitation [5] and heuristically optical path scanning (HOPS) [56, 73]. We further developed the HOPS approach to perform recordings in brain areas where neurons are sparsely distributed. Based on the assumption that most FOV pixels do not carry relevant signals (because of the sparseness of neurons), we introduced a software (developed in Matlab) that allows to extract a trajectory for a linescan acquisition of two-photon calcium imaging data. The trajectory covers and connects only the FOV areas where neurons are present (eventually, with a surround selected by the user), allowing to reduce the number of imaged pixels. We showed that switching from a raster scan acquisition to a linescan acquisition mode leads to improvements both in the temporal resolution and in the signal quality (measured using SNR) and allows to record data with single spike resolution. We validated in the LIII/IV of the somatosensory cortex of anesthetized and awake animals (abstract in Appendix C, paper under revision).

Given the great advantages provided by two-photon calcium imaging, we worked also on the second strong limit of this technique: imaging in deep brain areas. Two approaches are currently used for deep two-photon calcium imaging: optical windows or imaging with endoscopes. These approaches allows recording in deep brain areas, but present important drawbacks: optical windows are invasive and might destroy the neural circuits under investigation, while endoscopes preserve neural circuits but allow for imaging of small FOVs with a reduced number of neurons. We introduced and characterized a new class of aberration-corrected GRIN lens based microendoscopes, that combine small invasiveness with the chance of imaging a FOV of up to 500 μm * 500 μm. We showed through simulations the optical properties of this technique (larger FOV and better spatial resolution) and validated the use of the microendoscopes for the investigation of neural circuits by recording neural activity in the VPM nucleus of the thalamus of awake freely behaving mice (paper in preparation).

Application of these and similar techniques in experiments where animals perform a perceptual discrimination task, would allow to study with high spatial and temporal resolution the neural

code used for perception. Instead of splitting the investigation of perception in two stages, researchers might understand on a trial-to-trial basis the neural processes that lead from stimuli to behavior. An approach like the one presented in this work, that is the embedding of neural activity in a metric space, could theoretically be applied also to this kind of data and to the study of the code used to process the so-called *intersection information*, that is the information contained in neural activity about stimuli that is also used to inform behavior. In a very general way, a metric that reflects the mechanisms used to process intersection information, should be mapped both to stimuli (as in [74]) and to behavior (as in the current work) simultaneously. A further step in the understanding of the neural code is provided by the combination af this approach with interventional techniques. In this way, researcher would be able to draw conclusions not only about the correlation between stimuli, neural activity and behavior, but also on the causality of the relationships between these three elements.

Another future direction of this work that we are considering consists in the application of the metric framework to different (non-neural) signals. In particular, in collaboration with the group of Cristina Becchio at IIT in Genova, we would like to investigate the perceptual relevance of movement kinematics. In this case, we would build a metric based on different kinematic features and we would fit it to behavioral responses in order to investigate which features are more informative for behavior. This would provide an alternative approach to classical psychophysics techniques to study the link between sensory stimuli and behavior.

Appendix A

OB perceptual metric

In this Appendix, I describe the mathematical formulation of the metric used in Chapter 4 to describe the animal perceptual readout. In the first section, I describe the construction of the metric step-by-step. In the following section, I provide the proof of the properties that characterize a metric.

A.1 Metric definition

Let's assume that the neural space is a N-dimensional space, where N is the number of channels conveying neural signals (neurons, or glomeruli, or activated spots). We represent a spatio-temporal pattern of neural activity P as a $(N * 2)$-dimensional array:

$$P = \{(x_1, t_1), (x_2, t_2), ..., (x_N, t_N)\} \tag{A.1}$$

Where x_i is a binary variable indicating whether the i-th channel is active and t_i is the activation time of the i-th channel. If the i-th channel is not active, then we set t_i to 0 (without any loss of generality).

Given two patterns P and Q, we want to define a metric function, that is a function that maps

each pair of patterns P and Q into a non-negative value

$$d : (P, Q) \in \mathbb{R}^N * \mathbb{R}^N \rightarrow \mathbb{R}^+ \tag{A.2}$$

and satisfies the following conditions:

1. Non-negativity: $d(P, Q) \geq 0, \quad \forall P, Q$
2. Identity of indiscernibles: $d(P, Q) = 0 \iff P = Q, \quad \forall P, Q$
3. Symmetry: $d(P, Q) = d(Q, P), \quad \forall P, Q$
4. Triangular inequality: $d(P, Q) \leq d(P, R) + d(R, Q), \quad \forall P, Q, R$

A.1.1 From discrete signals to continuous waveforms

Given a spatio-temporal pattern, that is defined as a discrete object, we map it into a set of continous waveforms, applying the convolution with an exponential decaying kernel $h(t; \tau_{act})$ defined as:

$$h(t; \tau_{act}) = \begin{cases} e^{-\frac{t}{\tau_{act}}}, & \text{if } t > 0 \\ 0, & \text{otherwise} \end{cases} \tag{A.3}$$

with $\tau_{act} > 0$.

The mapping to continuous functions through convolution is defined as:

$$P = \{(x_1, t_1), (x_2, t_2), ..., (x_N, t_N)\} \mapsto \boldsymbol{\phi}^P(t; \tau_{primacy}, \tau_{act}) = \begin{cases} \phi_1^P(t; \tau_{primacy}, \tau_{act}) \\ \phi_2^P(t; \tau_{primacy}, \tau_{act}) \\ ... \\ \phi_N^P(t; \tau_{primacy}, \tau_{act}) \end{cases} \tag{A.4}$$

where

$$\phi_i^P(t;\tau_{primacy},\tau_{act}) = \begin{cases} e^{-\frac{t_i-t_1}{\tau_{primacy}}} * h(t-t_i), & \text{if } x_i = 1 \\ 0, & \text{otherwise} \end{cases} \tag{A.5}$$

and $\tau_{primacy} > 0$.

Given an angle $\theta \in [0, \frac{\pi}{2}]$, we consider in the N-dimensional space N axes rotated by an angle θ. We denote with $\boldsymbol{e}_i(\theta)$ the direction of the i-th axis and project the components of the function $\boldsymbol{\phi^P}$ on these axes:

$$P \mapsto \boldsymbol{f}^P(t;\tau_{primacy},\tau_{act},\theta) = \sum_{i=1}^{N} \phi_i^P(t;\tau_{primacy},\tau_{act})\boldsymbol{e}_i(\theta) \tag{A.6}$$

The angle θ can be interpreted as a parameter that represents the relevance of the channel identity in the readout. If $\theta = 0$, than the signals from all the channels are summed up "before patterns comparison" and channel identity is not relevant for behavioral readout. If $\theta = \frac{\pi}{2}$, then each channel is considered separately during patterns comparisons.

A.1.2 Definition of a center of activity

Once the patterns are represented through continuous functions, we we assign to each pattern P its *center of activity* (CA), which can be defined in different ways:

1. Timing of earliest activated spot: $CA_P = \min_i t_i$
2. Average of all spots activation: $CA_P = \frac{1}{\sum_{i=1}^{N} x_i} \sum_{x_i \neq 0} t_i$
3. Center of mass of the summed population activity:

$$CA_P = \frac{\int_{-\infty}^{+\infty} t f_1^P(t;\tau_{prim},\tau_{act},\theta=0)dt}{\int_{-\infty}^{+\infty} f_1^P(t;\tau_{prim},\tau_{act},\theta=0)dt} \tag{A.7}$$

4. Center of mass of the summed single channels cross-correlograms:

$$CA_P = \frac{\int_{-\infty}^{+\infty} \tau \sum_{i=1}^{N} CC_i(\tau;\tau_{prim},\tau_{act},\theta)d\tau}{\int_{-\infty}^{+\infty} \sum_{i=1}^{N} CC_i(\tau;\tau_{prim},\tau_{act},\theta)d\tau} \tag{A.8}$$

where CC_i denotes the cross-correlogram of the i-th channel signal:

$$CC_i(\tau;\tau_{prim},\tau_{act},\theta)=\int_{-\infty}^{+\infty}[\phi_i^P(t;\tau_{primacy},\tau_{act})\boldsymbol{e}_i(\theta)][\phi_i^P(t+\tau;\tau_{primacy},\tau_{act})\boldsymbol{e}_i(\theta)]dt \tag{A.9}$$

We use the center of activity to account for the position of a pattern in the sniff cycle. The choice of the center of activity depends on the properties of the system we are studying, for example on feedback inhibition processes or effects such as primacy.

A.1.3 Perceptual metric

Given two patterns P and Q, we define the perceptual metric function in the following steps:

1. Associate to each pattern its center of activity;

2. Compute the difference in global positioning within the sniff cycle, which is a function of the difference between the centers of activity of the two patterns: $d_{CA}=|CA_P-CA_Q|$.

3. Align the patterns to their center of activity. I denote with P' and Q' the shifted patterns.

$$P'=\{(x_1,t_1-CA_P),(x_2,t_2-CA_P),...,(x_N,t_N-CA_P)\} \tag{A.10}$$

4. The distance between the two patterns is given by:

$$\begin{aligned}&d(P,Q;\tau_{primacy},\tau_{act},\lambda_{glob},\tau_{glob},\theta)=\\&=d_{rel}(P',Q';\tau_{primacy},\tau_{act},\theta)+\lambda_{glob}(1-e^{-\frac{|CA_P-CA_Q|}{\tau_{glob}}})=\\&=d_{rel}(P',Q';\tau_{primacy},\tau_{act},\theta)+d_{glob}(CA_P,CA_Q;\lambda_{glob},\tau_{glob})\end{aligned} \tag{A.11}$$

where

$$\begin{aligned} & d_{rel}(P,Q;\tau_{primacy},\tau_{act},\theta) = \\ & = \|\boldsymbol{f}^P(t;\tau_{primacy},\tau_{act},\theta) - \boldsymbol{f}^Q(t;\tau_{primacy},\tau_{act},\theta)\|_2 \\ & = \sqrt{\sum_1^N \int_{-\infty}^{\infty} |f_i^P(t;\tau_{primacy},\tau_{act},\theta) - f_i^Q(t;\tau_{primacy},\tau_{act},\theta)|^2 dt} \end{aligned} \tag{A.12}$$

with $\lambda_{glob}, \tau_{glob} > 0$. We denote the first distance component as $d_{rel}(P,Q;\tau_{primacy},\tau_{act},\theta)$ because it provides a measure of the difference of the "relative" timing structure of the activation patterns, after having removed the pattern position in the sniff cycle.

A.2 Proof of metric properties

Before showing that the function $d(P,Q;\tau_{primacy},\tau_{act},\lambda_{glob},\tau_{glob},\theta)$ is a metric, we show some results about $d_{rel}(P,Q;\tau_{primacy},\tau_{act},\theta)$. For easiness we will denote $d_{rel}(P,Q;\tau_{primacy},\tau_{act},\theta)$ as $d_{rel}(P,Q)$ considering the dependence from the parameters implicit. The same for the other functions defined previously, unless parameters are necessary to show the results.

Lemma A.1. *Given any two patterns P and Q, the function $d_{rel}(P,Q;\tau_{primacy},\tau_{act},\theta)$ defined in A.12 satisfies the properties in section A.1 for any $\tau_{primacy},\tau_{act} \geq 0$ and $\theta \in (0,\frac{\pi}{2}]$. In other words, $d_{rel}(P,Q)$ is a metric.*

Proof. We need to show that all the metric properties are satisfied.

1. Non-negativity: $d_{rel}(P,Q) \geq 0, \quad \forall P,Q$

   $d_{rel}(P,Q) = \|\boldsymbol{f}^P(t) - \boldsymbol{f}^Q(t)\|_2 \geq 0, \quad \forall P,Q$

   for the properties of the L_2 norm.

2. Identity of indiscernibles: $d_{rel}(P,Q) = 0 \iff P = Q, \quad \forall P,Q$

   $d_{rel}(P,Q) = \|\boldsymbol{f}^P(t) - \boldsymbol{f}^Q(t)\|_2 = 0 \iff$

   $\iff f_i^P(t) - f_i^Q(t) = 0 \quad \forall i$ and almost everywhere in $t \iff$

$\iff \phi_i^P(t) = \phi_i^Q(t) \quad \forall i$ and almost everywhere in $t \iff P = Q$

for the definition of the L_2 norm and because $\phi_i^P(t)$ is a continuous function in t, $\forall i$ and $\forall P$.

3. Symmetry: $d_{rel}(P, Q) = d_{rel}(Q, P), \quad \forall P, Q$

   $d_{rel}(P, Q) = \|\boldsymbol{f}^P(t) - \boldsymbol{f}^Q(t)\|_2 = \|\boldsymbol{f}^Q(t) - \boldsymbol{f}^P(t)\|_2 = d_{rel}(Q, P), \quad \forall P, Q$

   for the properties of the L_2 norm.

4. Triangular inequality: $d_{rel}(P, Q) \leq d_{rel}(P, R) + d_{rel}(R, Q), \quad \forall P, Q, R$

   $d_{rel}(P, Q) = \|\boldsymbol{f}^P(t) - \boldsymbol{f}^Q(t)\|_2 \leq$

   $\leq \|\boldsymbol{f}^P(t) - \boldsymbol{f}^R(t)\|_2 + \|\boldsymbol{f}^R(t) - \boldsymbol{f}^Q(t)\|_2 = d_{rel}(P, R) + d_{rel}(R, Q), \quad \forall P, Q$

   for the properties of the L_2 norm.

□

Theorem A.1. *Given any two patterns P and Q, the function $d(P, Q; \tau_{primacy}, \tau_{act}, \lambda_{glob}, \tau_{glob}, \theta)$ defined in A.11 satisfies the properties in section A.1 for any $\tau_{primacy}, \tau_{act}, \lambda_{glob}, \tau_{glob} \geq 0$ and $\theta \in (0, \frac{\pi}{2}]$. In other words, $d(P, Q)$ is a metric.*

Proof. We need to show that all the metric properties are satisfied.

1. Non-negativity: $d(P, Q) \geq 0, \quad \forall P, Q$

   $d(P, Q) = d_{rel}(P', Q') + d_{glob}(CA_P, CA_Q) =$

   $= d_{rel}(P', Q') + \lambda_{glob}(1 - e^{-\frac{|CA_P - CA_Q|}{\tau_{glob}}}) \geq 0, \quad \forall P, Q$

   for the properties of d_{rel} shown in Lemma A.1 and the properties of the exponential function.

2. Identity of indiscernibles: $d(P, Q) = 0 \iff P = Q, \quad \forall P, Q$

   $d(P, Q) = d_{rel}(P', Q') + \lambda_{glob}(1 - e^{-\frac{|CA_P - CA_Q|}{\tau_{glob}}}) = 0 \iff$

$$\iff \begin{cases} d_{rel}(P', Q') = 0 \\ \lambda_{glob}(1 - e^{-\frac{|CA_P - CA_Q|}{\tau_{glob}}}) = 0 \end{cases} \iff \begin{cases} P' = Q' \\ e^{-\frac{|CA_P - CA_Q|}{\tau_{glob}}} = 1 \end{cases} \iff$$

$$\iff \begin{cases} P - CA_P = Q - CA_Q \\ |CA_P - CA_Q| = 0 \end{cases} \iff \begin{cases} P - CA_P = Q - CA_Q \\ CA_P = CA_Q \end{cases} \iff P = Q$$

for the properties of d_{rel} shown in Lemma A.1 and the properties of the exponential function.

3. Symmetry: $d(P, Q) = d(Q, P), \quad \forall P, Q$

   $d(P, Q) = d_{rel}(P', Q') + \lambda_{glob}(1 - e^{-\frac{|CA_P - CA_Q|}{\tau_{glob}}}) =$

   $= d_{rel}(Q', P') + \lambda_{glob}(1 - e^{-\frac{|CA_Q - CA_P|}{\tau_{glob}}}) = d(Q, P)$

   for the properties of d_{rel} shown in Lemma A.1 and the simmetry of the absolute value.

4. Triangular inequality: $d(P, Q) \leq d(P, R) + d(R, Q), \quad \forall P, Q, R$

   From the monotonicity of the function $f(x) = 1 - e^{-\frac{x}{\tau}}$ and from the absolute value properties

$$\lambda_{glob}(1 - e^{-\frac{|CA_P - CA_Q|}{\tau_{glob}}}) \leq \lambda_{glob}(1 - e^{-\frac{|CA_P - CA_R|}{\tau_{glob}}}) + \lambda_{glob}(1 - e^{-\frac{|CA_R - CA_Q|}{\tau_{glob}}}) \tag{A.13}$$

   Using Equation A.13 and Lemma A.1 we show now that the triangular inequality holds for $d(P, Q)$:

   $d(P, Q) = d_{rel}(P', Q') + \lambda_{glob}(1 - e^{-\frac{|CA_P - CA_Q|}{\tau_{glob}}}) \leq$

   $d_{rel}(P', R') + d_{rel}(R', Q') + \lambda_{glob}(1 - e^{-\frac{|CA_P - CA_R|}{\tau_{glob}}}) + \lambda_{glob}(1 - e^{-\frac{|CA_R - CA_Q|}{\tau_{glob}}}) =$

   $= d(P, R) + d(R, Q)$

□

Appendix B

Supplementary Figures

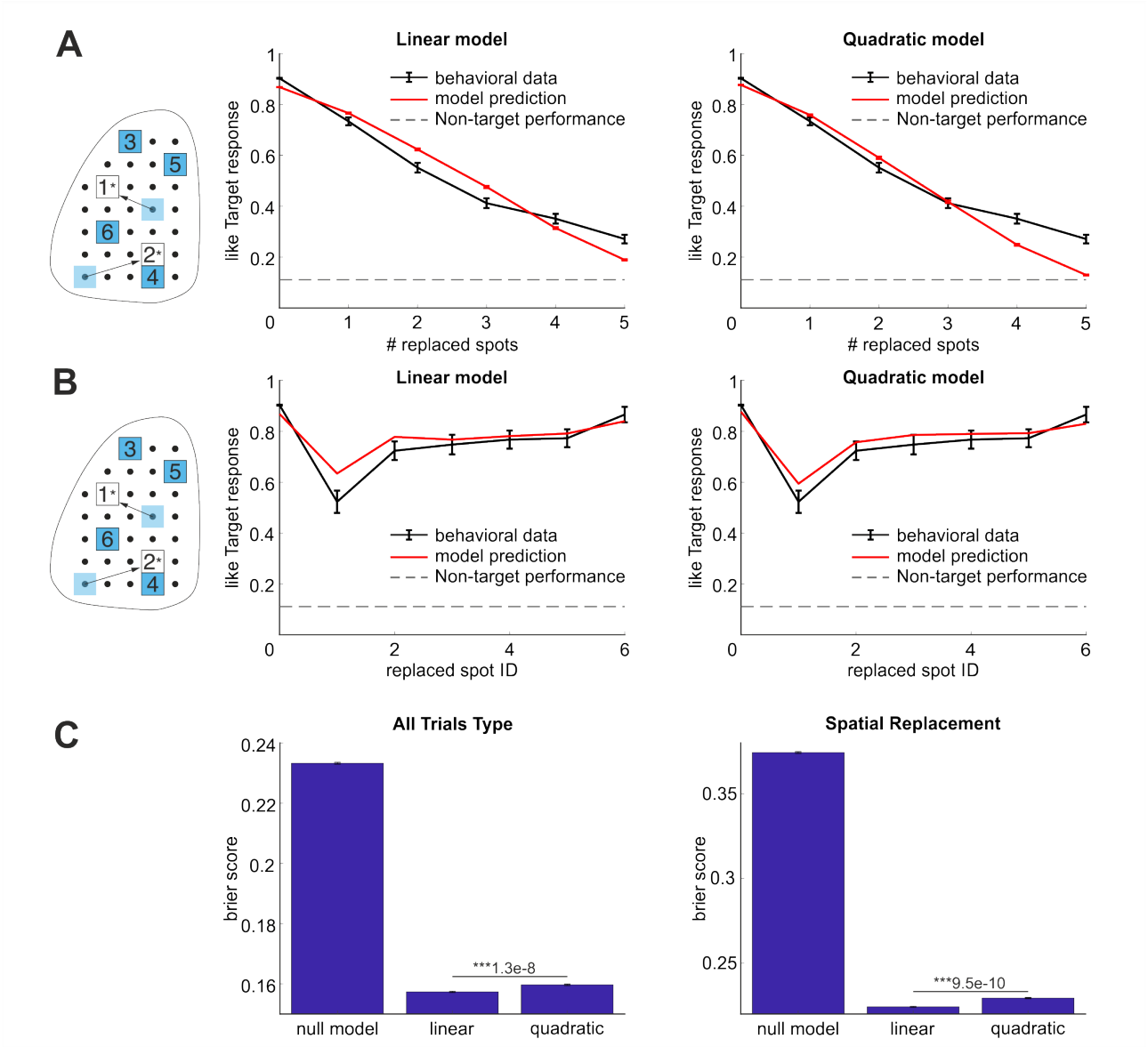

Figure B.1: **Logistic regression models for spatial perturbations. A.** Left. Illustration of spatial replacement trials. Middle and right. Effect on animal behavior of the replacement of a different number of spots. In (**A**) and (**B**) behavioral data (black) are plotted together with the predictions of the linear logistic regression model (red, middle) or of the quadratic logistic regression model (red, right). The dashed gray line shows the behavioral performance for Non-target trials. **B.** Left. Illustration of spatial replacement trials. Middle and right. Effect on animal behavior of identity of the replaced spot for single spot replacement trials. **C.** Prediction accuracy of the logistic regression models, measures as Brier Score on n=500 bootstrap samples of the test set. Left. Brier Score computed on all trials type (Target trial, Non-target trials andspatial replacement trials). Right. Brier Score computed on spatial replacement trials.

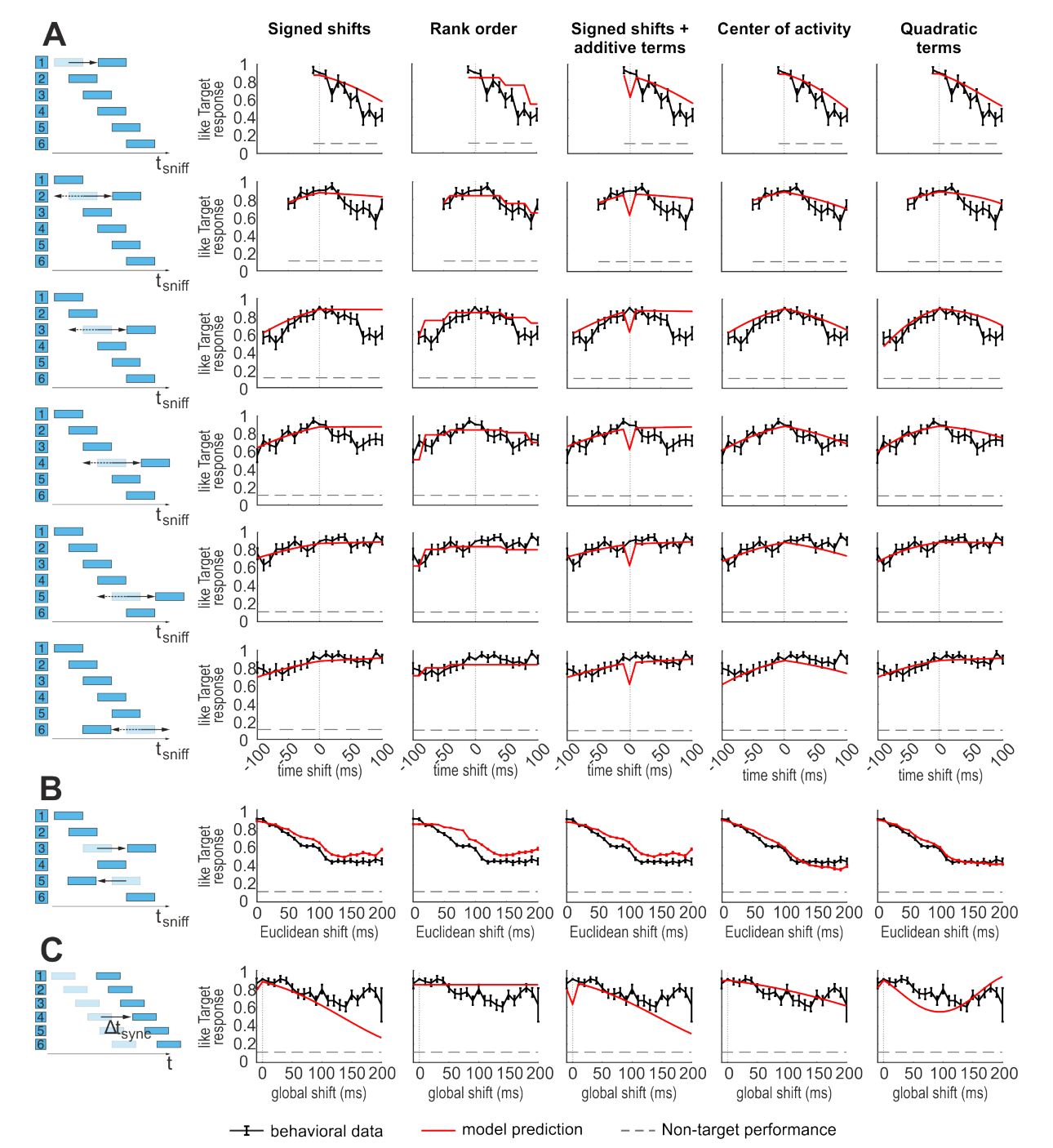

Figure B.2: **Logistic regression models for temporal perturbations.** In the following, figures in the left column illustrate the type of perturbation represented in the corresponding row. Each of the other columns shows the predictions obtained using different logistic regression models (from left to right: signed time shift model, rank order model, signed time shift model with additive interactions, center of activity model with relative shifts, quadratic model). **A.** Effect on animal behavior of the shift of each spot (only trials with single spot shift). **B.** Effect on animal behavior of multiple spots shift as a function of the total euclidean shift from Target pattern timing. **C.** Effect on animal behavior of the simultaneous shift of all spots. In (**A**), (**B**) and (**C**) behavioral data (black) are plotted together with the predictions of the considered logistic regression model (red). The dashed gray line shows the behavioral performance for Non-target trials.

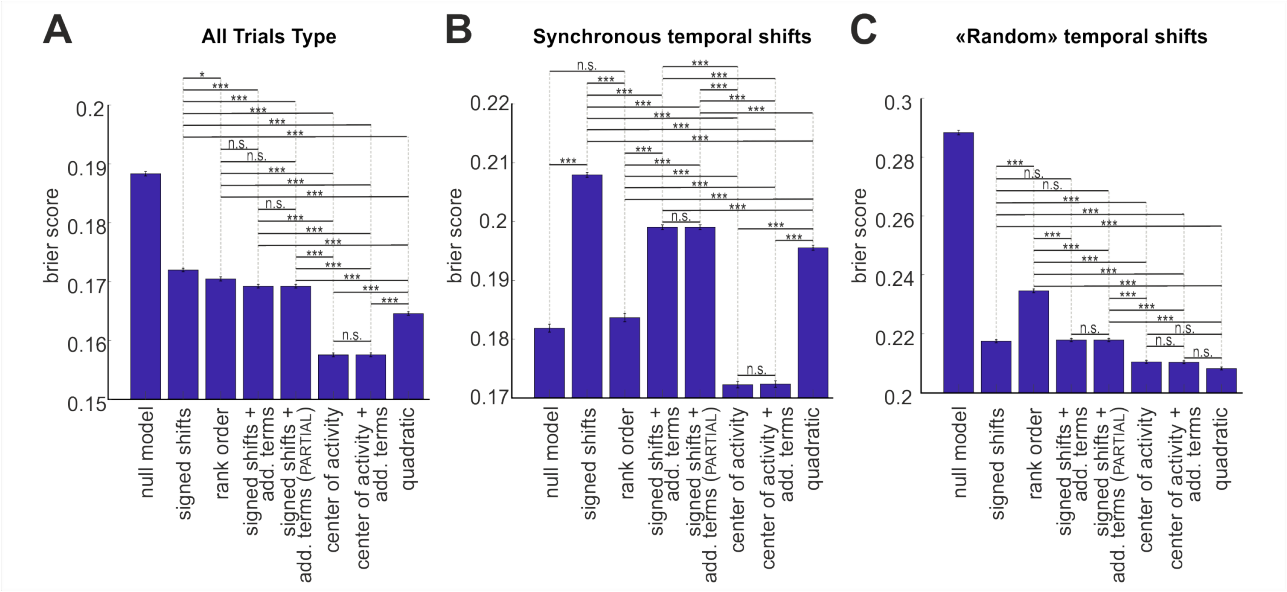

Figure B.3: **Logistic regression models for temporal perturbations. Brier score and models comparison. A.** Brier score computed for all the logistic regression models on all trials type (Target trial, Non-Target Trials, temporal synchronous perturbations and temporal "random" perturbations). **B.** Brier score computed for all the models on temporal synchronous trials. **C.** Brier score computed for all the logistic regression models on temporal "random" perturbations trials. In (**A**), (**B**) and (**C**) brier scores and p-values are computed on n=500 bootstrap samples of the test set (with balanced number of trials for each trial type).

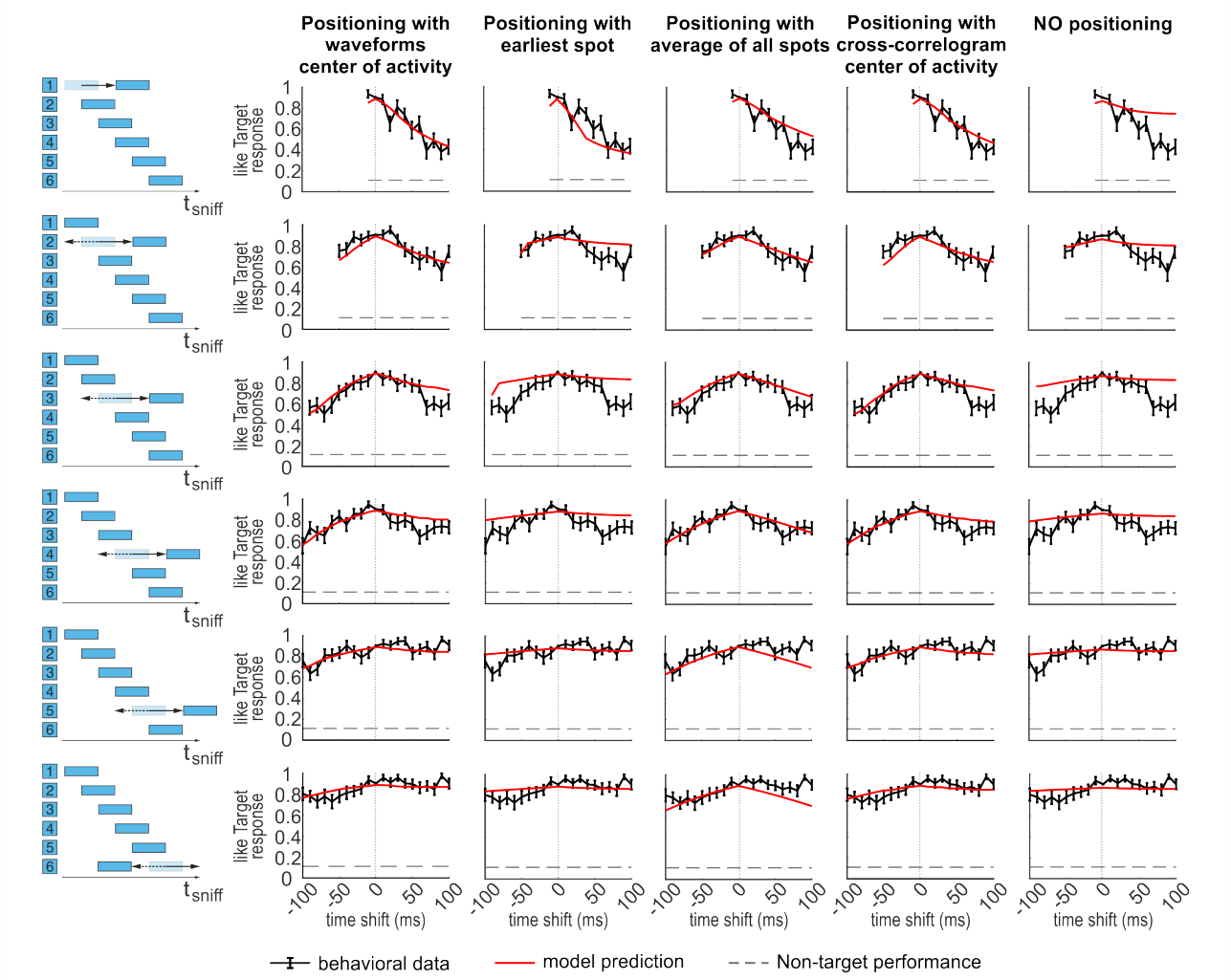

Figure B.4: **Perceptual metric. Pattern positioning in sniff cycle. Part 1.** In the following, figures in the left column illustrate the type of perturbation represented in the corresponding row. Each of the other columns shows the predictions obtained using different techniques to compute the pattern position in the sniff cycle (from left to right: center of activity, earliest activation timing, average activation timing, center of activity of the summed single-channel cross-correlograms, without positioning in the sniff cycle). Results are reported for single spot temporal perturbation trials. Behavioral data (black) are plotted together with the predictions of the perceptual metric with the considered sniff-positioning technique (red). The dashed gray line shows the behavioral performance for Non-target trials.

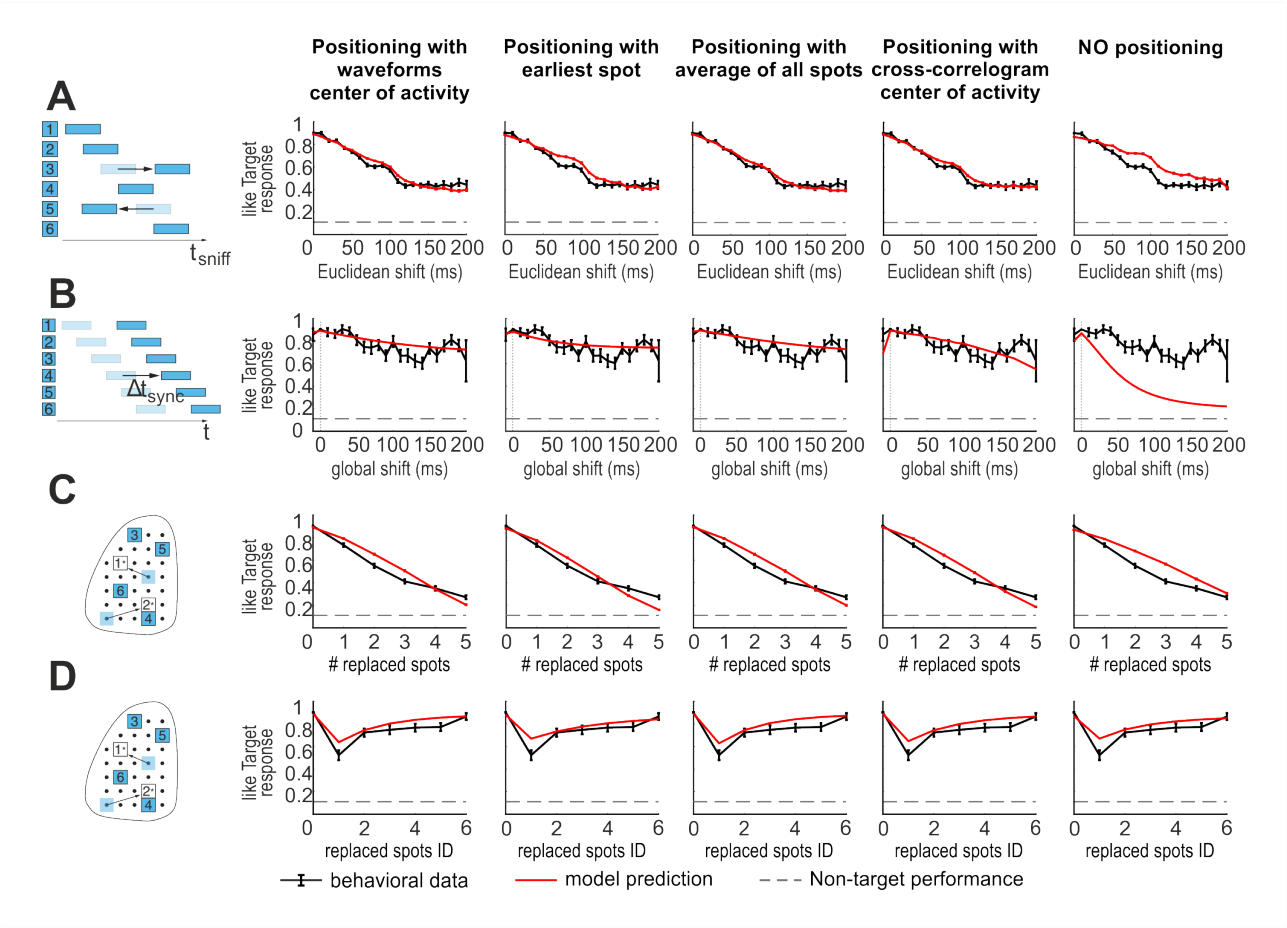

Figure B.5: **Perceptual metric. Pattern positioning in sniff cycle. Part 2.** In the following, figures in the left column illustrate the type of perturbation represented in the corresponding row. Each of the other columns shows the predictions obtained using different techniques to compute the pattern position in the sniff cycle (from left to right: center of activity, earliest activation timing, average activation timing, center of activity of the summed single-channel cross-correlograms, without positioning in the sniff cycle). **A.** Effect on animal behavior of multiple spots shift as a function of the total euclidean shift from Target pattern timing. **B.** Effect on animal behavior of the simultaneous shift of all spots. **C.** Effect on animal behavior of the replacement of a different number of spots. **D.** Effect on animal behavior of identity of the replaced spot for single spot replacement trials. In (**A**), (**B**), (**C**) and (**D**) behavioral data (black) are plotted together with the predictions of the perceptual metric with the considered sniff-positioning technique (red). The dashed gray line shows the behavioral performance for Non-target trials.

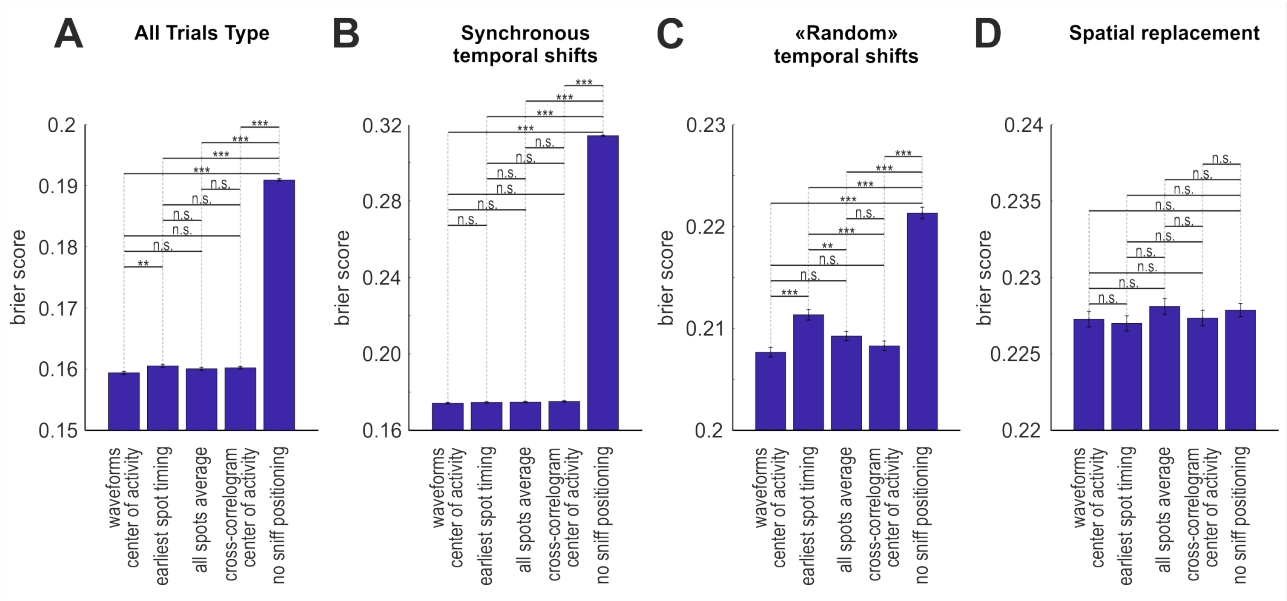

Figure B.6: **Perceptual metric: pattern position in the sniff cycle. Brier score and models comparison. A.** Brier score computed on all trials type (Target trial, Non-Target Trials, temporal synchronous perturbations, temporal "random" perturbations and spatial perturbations) for the different techniques to compute the position of patterns in the sniff-cycle for the perceptual metric definition. **B.** Brier score computed for all the sniff-positioning techniques on temporal synchronous trials. **C.** Brier score computed for all the sniff-positioning techniques on temporal "random" perturbations trials. **D.** Brier score computed for all the sniff-positioning techniques on spatial perturbations trials. In (**A**), (**B**), (**C**) and (**D**) brier scores and p-values are computed on n=500 bootstrap samples of the test set (with balanced number of trials for each trial type).

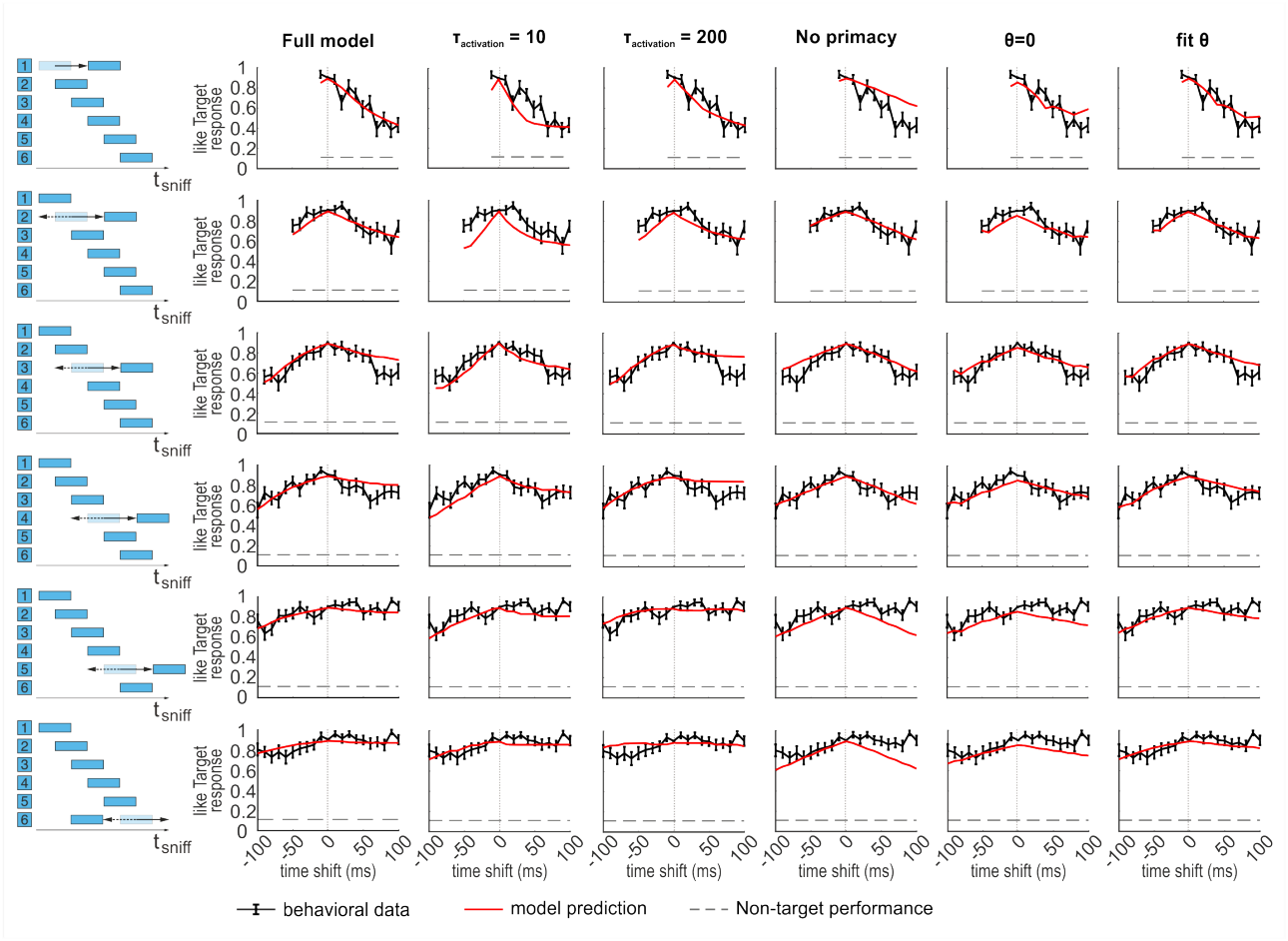

Figure B.7: **Perceptual metric: remove components. Part 1. A.** In the following, figures in the left column illustrate the type of perturbation represented in the corresponding row. Each of the other columns shows the predictions obtained when single components were removed from the perceptual metric definition (from left to right: full model, $tau_{activation} = 10$, $tau_{activation} = 200$, without primacy modulation, $\theta = 0$, fit of θ.). Results are reported for single spot temporal perturbation trials. Behavioral data (black) are plotted together with the predictions of the perceptual metric without the considered component (red). The dashed gray line shows the behavioral performance for Non-target trials.

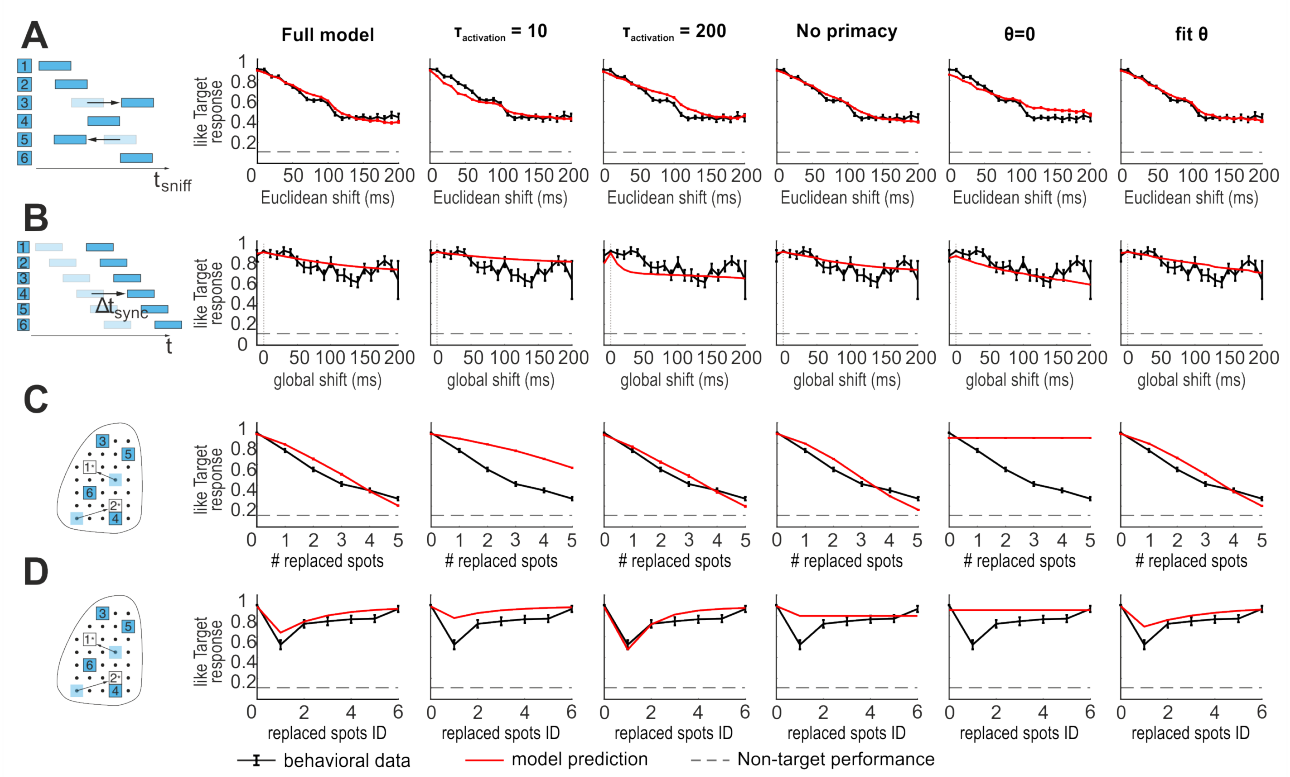

Figure B.8: **Perceptual metric: remove components. Part 2.** In the following, figures in the left column illustrate the type of perturbation represented in the corresponding row.Each of the other columns shows the predictions obtained when single components were removed from the perceptual metric definition (from left to right: full model, $tau_{activation} = 10$, $tau_{activation} = 200$, without primacy modulation, $\theta = 0$, fit of θ.). **A.** Effect on animal behavior of multiple spots shift as a function of the total euclidean shift from Target pattern timing. **B.** Effect on animal behavior of the simultaneous shift of all spots. **C.** Effect on animal behavior of the replacement of a different number of spots. **D.** Effect on animal behavior of identity of the replaced spot for single spot replacement trials. In (**A**), (**B**), (**C**) and (**D**) behavioral data (black) are plotted together with the predictions of the perceptual metric without the considered component (red). The dashed gray line shows the behavioral performance for Non-target trials.

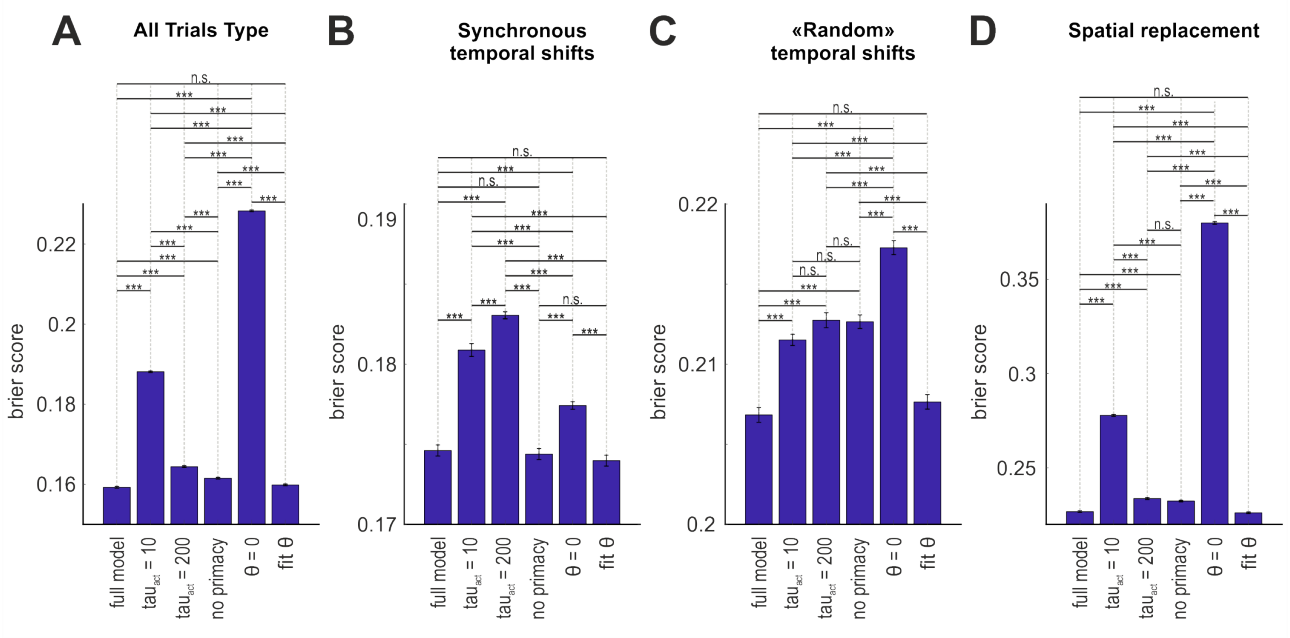

Figure B.9: **Perceptual metric: remove components. Brier score and models comparison.** **A.** Brier score computed on all trials type (Target trial, Non-Target Trials, temporal synchronous perturbations, temporal "random" perturbations and spatial perturbations) for the perceptual metric definitions where single components were removed. **B.** Brier score computed on temporal synchronous trials for all the component-lacking versions of the perceptual metric. **C.** Brier score computed on temporal "random" perturbations trials for all the component-lacking versions of the perceptual metric. **D.** Brier score computed on spatial perturbations trials for all the component-lacking versions of the perceptual metric. In (**A**), (**B**), (**C**) and (**D**) brier scores and p-values are computed on n=500 bootstrap samples of the test set (with balanced number of trials for each trial type).

Appendix C

Abstracts

Spatio-temporal matching of neural activity explains olfactory perception

Edmund Chong, Monica Moroni, Christopher Wilson, Shy Shoham, Stefano Panzeri, Dmitry Rinberg

Precise quantification of how spatio-temporal structures in neural activity are consequential for behavior, remains challenging. We developed a novel experimental approach and theoretical framework within mouse olfaction, where odors evoke dynamic, spatio-temporal olfactory bulb activity. We trained mice to recognize synthetic odors constructed from precise, parametrically - defined optogenetic stimulation patterns. We then measured perceptual changes during controlled perturbations across various spatial and temporal dimensions, achieving manipulation breadth and precision previously prohibitive using conventional odors. We found that animals responses could be modelled by spatio-temporal matching of bulb activation, where spatial components combine linearly, latencies of activation are defined relative to the entire pattern, and early activated components are preferentially weighted. Our synthetic approach reveals the fundamental logic of the olfactory code, and provides a novel framework for testing the link between spatio-temporal neural codes and behavior.

High accuracy detection of neural ensemble activity in two-photon functional microscopy using smart line scanning

Marco Brondi, Monica Moroni, Dania Vecchia, Manuel Molano-Mazon, Stefano Panzeri, Tommaso Fellin

Two-photon functional imaging using the genetically encoded calcium indicators (GECIs) represents one preferred tool to map neural activity. Under optimized experimental conditions GECIs detect single action potentials in individual cells with high accuracy. However, using current approaches these optimized conditions are never met when imaging large ensembles of neurons. Here, we developed a method that substantially increases the signal-to-noise ratio (SNR) of population imaging of GECIs using standard galvanometric mirrors and fast smart line scan trajectories. We validated our approach in anesthetized and awake mice on deep and dense GCaMP6 staining in the mouse barrel cortex during spontaneous and sensory-evoked activities. Compared to raster population imaging, smart line scan led to increased SNR, higher probability of detecting calcium events, and more precise identification of functional neural ensembles. Smart line scan provides a cheap and easily implementable tool for high accuracy population imaging of GCaMP6s signals using standard galvanometric-based two-photon microscopes.

Bibliography

[1] A. Akrami, C. Kopec, M. E. Diamond, and C. D. Brody. Posterior parietal cortex represents sensory stimulus history and is necessary for its effects on behavior. *Nature*, 02 2018.

[2] A. Apicella, Q. Yuan, M. Scanziani, and J. S. Isaacson. Pyramidal cells in piriform cortex receive convergent input from distinct olfactory bulb glomeruli. *Journal of Neuroscience*, 30(42):14255–14260, 2010.

[3] D. Aronov, D. S. Reich, F. Mechler, and J. D. Victor. Neural coding of spatial phase in v1 of the macaque monkey. *Journal of Neurophysiology*, 89(6):3304–3327, 2003. PMID: 12612048.

[4] K. A. Bolding and K. M. Franks. Complementary codes for odor identity and intensity in olfactory cortex. *eLife*, 6:e22630, apr 2017.

[5] S. Bovetti, C. Moretti, S. Zucca, M. Dal Maschio, P. Bonifazi, and T. Fellin. Simultaneous high-speed imaging and optogenetic inhibition in the intact mouse brain. *Scientific Reports*, 7:40041, 01 2017.

[6] T. Bozza, P. Feinstein, C. Zheng, and P. Mombaerts. Odorant receptor expression defines functional units in the mouse olfactory system. *Journal of Neuroscience*, 22(8):3033–3043, 2002.

[7] K. H. Britten, W. T. Newsome, M. N. Shadlen, S. Celebrini, and J. A. Movshon. A relationship between behavioral choice and the visual responses of neurons in macaque mt. *Visual Neuroscience*, 13(1):87100, 1996.

[8] V. Bruce, P. R. Green, and M. Georgeson. *Visual perception : Physiology, psychology and ecology*. Psychology Press, 4th edition, 2003.

[9] L. Busse, A. Ayaz, N. T. Dhruv, S. Katzner, A. B. Saleem, M. L. Schölvinck, A. D. Zaharia, and M. Carandini. The detection of visual contrast in the behaving mouse. *Journal of Neuroscience*, 31(31):11351–11361, 2011.

[10] G. Buzsáki. Neural syntax: Cell assemblies, synapsembles, and readers. *Neuron*, 68(3):362 – 385, 2010.

[11] R. M. Carey, J. V. Verhagen, D. W. Wesson, N. Pírez, and M. Wachowiak. Temporal structure of receptor neuron input to the olfactory bulb imaged in behaving rats. *Journal of Neurophysiology*, 101(2):1073–1088, 2009. PMID: 19091924.

[12] T. W. Chen, T. Wardill, Y. Sun, S. R. Pulver, S. L. Renninger, A. Baohan, E. R. Schreiter, R. Kerr, M. Orger, V. Jayaraman, L. L. Looger, K. Svoboda, and D. S. Kim. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, 499:295–300, 07 2013.

[13] M. R. Cohen and J. H. R. Maunsell. Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, 12(12):1594–1600, Nov 2009.

[14] T. B. Crapse and M. A. Basso. Insights into decision making using choice probability. *Journal of Neurophysiology*, 114(6):3039–3049, 2015. PMID: 26378203.

[15] V. de Lafuente and R. Romo. Neural correlate of subjective sensory experience gradually builds up across cortical areas. *Proceedings of the National Academy of Sciences*, 103(39):14266–14271, 2006.

[16] A. K. Dhawal, A. Hagiwara, U. S. Bhalla, V. N. Murthy, and D. F. Albeanu. Non-redundant odor coding by sister mitral cells revealed by light addressable glomeruli in the mouse. *Nature Neuroscience*, 13:1404–1412, 2010.

[17] A. Forli, D. Vecchia, N. Binini, F. Succol, S. Bovetti, C. Moretti, F. Nespoli, M. Mahn, C. A. Baker, M. M. Bolton, O. Yizhar, and T. Fellin. Two-photon bidirectional control

and imaging of neuronal excitability with high spatial resolution in vivo. *Cell Reports*, 22(11):3087 – 3098, 2018.

[18] K. M. Franks and J. S. Isaacson. Strong single-fiber sensory inputs to olfactory cortex: Implications for olfactory coding. *Neuron*, 49(3):357 – 363, 2006.

[19] B. F. Grewe, D. Langer, H. Kasper, B. M. Kampa, and F. Helmchen. High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision. *Nature Methods*, 7, 2010.

[20] C. Grienberger and A. Konnerth. Imaging calcium in neurons. *Neuron*, 73:862–85, 03 2012.

[21] Y. Gu, S. Liu, C. R. Fetsch, Y. Yang, S. Fok, A. Sunkara, G.C. DeAngelis, and D. E. Angelaki. Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron*, 71(4):750 – 761, 2011.

[22] Z. V. Guo, N. Li, D. Huber, E. Ophir, D. Gutnisky, J. T. Ting, G. Feng, and K. Svoboda. Flow of cortical activity underlying a tactile decision in mice. *Neuron*, 81(1):179 – 194, 2014.

[23] P. Gupta, D. F. Albeanu, and S. U. Bhalla. Olfactory bulb coding of odors, mixtures and sniffs is a linear sum of odor time profiles. *Nature neuroscience*, 18, 01 2015.

[24] R. Haddad, A. Lanjuin, L. Madisen, H. Zeng, V. Murthy, and N. Uchida. Olfactory cortical neurons read out a relative time code in the olfactory bulb. *Nature neuroscience*, 16, 05 2013.

[25] R. M. Haefner, S. Gerwinn, J. Macke, and M. Bethge. Inferring decoding strategies from choice probabilities in the presence of correlated variability. *Nature neuroscience*, 16, 01 2013.

[26] A. Hernández, V. Nácher, R. Luna, A. Zainos, L. Lemus, M. Alvarez, Y. Vázquez, L. Camarillo, and R. Romo. Decoding a perceptual decision process across cortex. *Neuron*, 66(2):300 – 314, 2010.

[27] A. E. Hoerl and Kennard R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[28] J. Hopfield. Pattern recognition computation using action potential timing for stimulus representation. *Nature*, 376:33–6, 08 1995.

[29] C. Houghton and K. Sen. A new multineuron spike train metric. *Neural Computation*, 20(6):1495–1511, 2008.

[30] A. Houweling and M. Brecht. Behavioural report of single neuron stimulation in somatosensory cortex. *Nature*, 451:65–8, 02 2008.

[31] D. H. Hubel and T. N. Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148(3):574–591, 1959.

[32] R. Iwata, H. Kiyonari, and T. Imai. Mechanosensory-based phase coding of odor identity in the olfactory bulb. *Neuron*, 96(5):1139 – 1152.e7, 2017.

[33] M. Jazayeri and A. Afraz. Navigating the neural space in search of the neural code. *Neuron*, 93(5):1003 – 1014, 2017.

[34] A. L. Juavinett, J. C. Erlich, and Churchland. A. K. Decision-making behaviors: weighing ethology, complexity, and sensorimotor compatibility. *Current Opinion in Neurobiology*, 49:42 – 50, 2018. Neurobiology of Behavior.

[35] S. Junek, E. Kludt, F. Wolf, and D. Schild. Olfactory coding with patterns of response latencies. *Neuron*, 67(5):872 – 884, 2010.

[36] J. L. Katz, L. andYates, J. W. Pillow, and A. C. Huk. Dissociated functional significance of decision-related activity in the primate dorsal stream. *Nature*, 535, 07 2016.

[37] A. Koulakov, A. Gelperin, and D. Rinberg. Olfactory coding with all-or-nothing glomeruli. *Journal of Neurophysiology*, 98(6):3134–3142, 2007. PMID: 17855585.

[38] G. M. Lerman, J. V. Gill, D. Rinberg, and S. Shoham. Precise optical probing of perceptual detection. *bioRxiv*, 2018.

[39] A. M. Licata, M. T. Kaufman, D. Raposo, M. B. Ryan, J. P. Sheppard, and A. K. Churchland. Posterior parietal cortex guides visual decisions in rats. *Journal of Neuroscience*, 37(19):4954–4966, 2017.

[40] B. Malnic, J. Hirono, T. Sato, and L. B. Buck. Combinatorial receptor codes for odors. *Cell*, 96(5):713 – 723, 1999.

[41] A. Mathis, D. Rokni, V. Kapoor, M. Bethge, and V. N. Murthy. Reading out olfactory receptors: Feedforward circuits detect odors in mixtures without demixing. *Neuron*, 91(5):1110 – 1123, 2016.

[42] H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264:746–748, 12 1976.

[43] K. Mori, Y. K. Takahashi, K. M. Igarashi, and M. Yamaguchi. Maps of odorant molecular features in the mammalian olfactory bulb. *Physiological Reviews*, 86(2):409–433, 2006. PMID: 16601265.

[44] Katz L. N., Yates J. L., Pillow J. W., and Huk A. C. Dissociated functional significance of decision-related activity in the primate dorsal stream. *Nature*, 2016.

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

[46] H. Nienborg and B. G. Cumming. Decision-related activity in sensory neurons reflects more than a neurons causal effect. *Nature*, 459:89–92, 04 2009.

[47] R. Nogueira, N. E. Peltier, A. Anzai, G. C. DeAngelis, J. Martínez-Trujillo, and R. Moreno-Bote. Identifying the most influential features of neural population responses for information encoding and behavior. *bioRxiv*, 2019.

[48] A. M. Packer, L. E. Russell, H. W. P. Dalgleish, and M. Häusser. Simultaneous all-optical manipulation and recording of neural circuit activity with cellular resolution in vivo. *Nature Methods*, 12(2):140–146, Dec 2014.

[49] S. Panzeri, N. Brunel, N. K. Logothetis, and C. Kayser. Sensory neural codes using multiplexed temporal scales. *Trends in Neurosciences*, 33(3):111 – 120, 2010.

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

[51] S. P. Peron, J. Freeman, V. Iyer, C. Guo, and K. Svoboda. A cellular resolution map of barrel cortex activity during tactile behavior. *Neuron*, 86(3):783 – 799, 2015.

[52] J. C. A. Read. The place of human psychophysics in modern neuroscience. *Neuroscience*, 296:116 – 129, 2015. Contributions from different model organisms to brain research.

[53] M. C. W. van Rossum. A novel spike distance. *Neural Computation*, 13(4):751–763, 2001.

[54] C. Runyan, E. Piasini, S. Panzeri, and C. Harvey. Distinct timescales of population coding across cortex. *Nature*, 548, 07 2017.

[55] N. C. Rust and J. A. Movshon. In praise of artifice. *Nature Neuroscience*, 8:1647, nov 2005.

[56] A. J. Sadovsky, P. B. Kruskal, J. M. Kimmel, J. Ostmeyer, F. B. Neubauer, and J. N. MacLean. Heuristically optimal path scanning for high-speed multiphoton circuit imaging. *Journal of Neurophysiology*, 106(3):1591–1598, 2011. PMID: 21715667.

[57] C. D. Salzman, C. M. Murasugi, K. H. Britten, and W. T. Newsome. Microstimulation in visual area mt: effects on direction discrimination performance. *Journal of Neuroscience*, 12(6):2331–2355, 1992.

[58] E. Schneidman. Towards the design principles of neural population codes. *Current Opinion in Neurobiology*, 37:133 – 140, 2016. Neurobiology of cognitive behavior.

[59] S. R. Schultz, C. S. Copeland, A. J. Foust, P. Quicke, and R. Schuck. Advances in two-photon scanning and scanless microscopy technologies for functional neural circuit imaging. *Proceedings of the IEEE*, 105(1):139–157, Jan 2017.

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

[61] R. N. Shepard. Toward a universal law of generalization for psychological science. *Science*, 237(4820):1317–1323, 1987.

[62] R. Shusterman, Y. B. Sirotin, M. C. Smear, Y. Ahmadian, and D. Rinberg. Sniff invariant odor coding. *eNeuro*, 5(6), 2018.

[63] R. Shusterman, M. Smear, A. Koulakov, and D. Rinberg. Precise olfactory responses tile the sniff cycle. *Nature neuroscience*, 14:1039–44, 07 2011.

[64] M. Smear, A. Resulaj, J. Zhang, T. Bozza, and D. Rinberg. Multiple perceptible signals from a single olfactory glomerulus. *Nature neuroscience*, 16, 09 2013.

[65] M. Smear, R. Shusterman, R. O'Connor, T. Bozza, and D. Rinberg. Perception of sniff phase in mouse olfaction. *Nature*, 479:397–400, 11 2011.

[66] Z. Soh, S. Nishikawa, Y. Kurita, N. Takiguchi, and T. Tsuji. A mathematical model of the olfactory bulb for the selective adaptation mechanism in the rodent olfactory system. *PLOS ONE*, 11(12):1–16, 12 2016.

[67] H. Spors, M. Wachowiak, L. B. Cohen, and R. W. Friedrich. Temporal dynamics and latency patterns of receptor neuron input to the olfactory bulb. *Journal of Neuroscience*, 26(4):1247–1259, 2006.

[68] S. M. Sternson and B. L. Roth. Chemogenetic tools to interrogate brain functions. *Annual Review of Neuroscience*, 37(1):387–407, 2014. PMID: 25002280.

[69] E. J. Tehovnik, A. S. Tolias, F. Sultan, W. M. Slocum, and N. K. Logothetis. Direct and indirect activation of cortical neurons by electrical microstimulation. *Journal of Neurophysiology*, 96(2):512–521, 2006. PMID: 16835359.

[70] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[71] G. Tkačik, E. Granot-Atedgi, R. Segev, and E. Schneidman. Retinal metric: A stimulus distance measure derived from population neural responses. *Phys. Rev. Lett.*, 110:058104, Jan 2013.

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

[73] I. Valmianski, A. Y. Shih, J. D. Driscoll, D. W. Matthews, Y. Freund, and D. Kleinfeld. Automatic identification of fluorescently labeled brain cells for rapid functional imaging. *Journal of Neurophysiology*, 104(3):1803–1811, 2010. PMID: 20610792.

[74] J. D. Victor and K. P. Purpura. Nature and precision of temporal coding in visual cortex: a metric-space analysis. *Journal of Neurophysiology*, 76(2):1310–1326, 1996. PMID: 8871238.

[75] J. D. Victor and K. P. Purpura. Metric-space analysis of spike trains: theory, algorithms and application. *Network: Computation in Neural Systems*, 8(2):127–164, 1997.

[76] Z. Williams, J. Elfar, E. Eskandar, L. J. Toth, and J. Assad. Parietal activity and the perceived direction of ambiguous apparent motion. *Nature neuroscience*, 6:616–23, 07 2003.

[77] C. D. Wilson, G. O. Serrano, A. Koulakov, and D. Rinberg. A primacy code for odor identity. *Nature Communications*, 8, 12 2017.

[78] Y. Yang, M. Deweese, G. H. Otazu, and A. Zador. Millisecond-scale differences in neural activity in auditory cortex can drive decisions. *Nature neuroscience*, 11:1262–3, 11 2008.

[79] Y. Yang and A. M. Zador. Differences in sensitivity to neural timing among cortical areas. *Journal of Neuroscience*, 32(43):15142–15147, 2012.