

# **Statistical Frameworks for Sensory Processing in the Mammalian Brain**

## **Dissertation**

der Mathematisch-Naturwissenschaftlichen Fakultät  
der Eberhard Karls Universität Tübingen  
zur Erlangung des Grades eines  
Doktors der Naturwissenschaften  
(Dr. rer. nat.)

vorgelegt von  
Lisa Schmors  
aus Potsdam

Tübingen  
2024



Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der  
Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:

25.10.2024

Dekan:

Prof. Dr. Thilo Stehle

1. Berichterstatter/-in:

Prof. Dr. Philipp Berens

2. Berichterstatter/-in:

Prof. Dr. Fabian Sinz

Copyright © 2024 by Lisa Schmors  
All Rights Reserved

### **Erklärung**

Ich, Lisa Schmors, erkläre, dass ich die zur Promotion eingereichte Arbeit mit dem Titel "*Statistical frameworks for sensory processing in the mammalian brain*", selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

### **Declaration**

I, Lisa Schmors, hereby declare that I have produced the work entitled "*Statistical frameworks for sensory processing in the mammalian brain*", submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

---

*Date*

---

*Signature*



Mind matters most.

— S. N. Goenka



# Abstract

Our perception of the world and all our actions are determined by our senses. This sensory information, represented by neuronal activity, is modulated by both behavioral and arousal state.

In neuroscience, a wide range of phenomenological models such as descriptive statistical methods, dimensionality reduction techniques, and generalized linear models, are used to find patterns and make predictions. They explain relationships between relevant variables and allow us to discover underlying mechanisms.

In this thesis, I discuss four studies involving neurons recorded in the mouse brain during spontaneous activity and in response to external stimuli. The first study investigates feedforward and feedback processing in the cortico-thalamic pathway using a data-efficient version of a generalized linear model. The second study uses unsupervised clustering of recordings in the lateral habenula to reveal cellular heterogeneity and its role in processing aversive information. The third study focuses specifically on the neuronal mechanisms underlying threat association in the lateral habenula and the dynamic neural encoding during cue-punishment learning. Lastly, the fourth study classifies neurons in the locus coeruleus based on distinct electrophysiological properties into noradrenergic and non-noradrenergic neurons, highlighting the complexity of neurotransmitter-specific neuronal dynamics in stress regulation.

The synergies between statistics and neuroscience hold vast potential for leveraging our understanding of the brain. At the same time, neuroscience can inspire more robust artificial systems and improve machine learning model interpretability. Taken together, this thesis presents an innovative example of research at the intersection of biology and data science, demonstrating how insights can be generated by applying statistical tools to neural data.

# Zusammenfassung

Unsere Wahrnehmung und alle unsere Handlungen werden durch unsere Sinne bestimmt. Die sensorische Information, die in neuronaler Aktivität kodiert ist, wird sowohl durch Verhaltens- als auch durch Erregungszustände beeinflusst. In den Neurowissenschaften werden eine Vielzahl von statistischen Modellen verwendet, um Daten zu analysieren und Vorhersagen zu treffen. Statistische Modelle erklären Beziehungen zwischen relevanten Variablen und ermöglichen es, zugrunde liegende Mechanismen zu verstehen.

Diese Dissertation enthält vier Studien, in denen die neuronale Aktivität von Mäusen untersucht wurde. Die erste Studie untersucht die Verarbeitung kortikaler Rückkopplungssignalen im Thalamus mithilfe eines generalisierten linearen Modells. Die zweite Studie verwendet einen Gruppierungsalgorithmus, um die zelluläre Heterogenität in der lateralen Habenula und die Verarbeitung aversiver Informationen zu untersuchen. Die dritte Studie erforscht genauer, welche neuronalen Mechanismen den Bedrohungsassoziationen in der Habenula zugrunde liegen. In der letzten Studie werden Neurone im Locus coeruleus basierend auf ihren unterschiedlichen elektrophysiologischen Eigenschaften in noradrenerge und nicht-noradrenerge Neuronen klassifiziert. Die Ergebnisse weisen auf neurotransmitter-spezifische Dynamiken bei der Stressregulation hin.

Die Synergien zwischen Statistik und Neurowissenschaften bieten enormes Potenzial. Erkenntnisse über neuronale Mechanismen können auch helfen die Robustheit von künstlichen Systemen und die Interpretierbarkeit von maschinellen Lernmodellen zu verbessern. Insgesamt stellt diese Dissertation ein innovatives Beispiel für Forschung an der Schnittstelle von Biologie und Datenwissenschaft dar und zeigt, wie wir durch die Anwendung von statistischen Analysen unser Gehirn besser verstehen können.

# Table of Contents

|                                                                                                                                                                       |              |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
| <b>Abstract</b>                                                                                                                                                       | <b>viii</b>  |
| <b>Table of Contents</b>                                                                                                                                              | <b>x</b>     |
| <b>List of Figures</b>                                                                                                                                                | <b>xv</b>    |
| <b>Notation</b>                                                                                                                                                       | <b>xviii</b> |
| <b>1 INTRODUCTION</b>                                                                                                                                                 | <b>1</b>     |
| <b>2 BACKGROUND</b>                                                                                                                                                   | <b>5</b>     |
| 2.1 Principles of neural processing . . . . .                                                                                                                         | 5            |
| 2.1.1 Sensory processing . . . . .                                                                                                                                    | 5            |
| 2.1.2 The computational significance of cell type diversity . . .                                                                                                     | 10           |
| 2.1.3 Contextual modulation of neural activity . . . . .                                                                                                              | 13           |
| 2.1.4 Plasticity and associative learning . . . . .                                                                                                                   | 16           |
| 2.2 Statistical frameworks for neural data . . . . .                                                                                                                  | 17           |
| 2.2.1 Linear and non-linear predictors . . . . .                                                                                                                      | 18           |
| 2.2.2 Decoders . . . . .                                                                                                                                              | 21           |
| 2.2.3 Clustering . . . . .                                                                                                                                            | 24           |
| 2.2.4 Visualizations in low dimensional space . . . . .                                                                                                               | 29           |
| <b>3 RESULTS</b>                                                                                                                                                      | <b>32</b>    |
| 3.1 <b>Study I:</b> The combination of stimulus-driven and modulatory in-<br>puts in visual thalamus depend on visual responsiveness and stim-<br>ulus type . . . . . | 33           |
| 3.1.1 dLGN responses to movies are modulated by behavioural<br>variables and CT feedback suppression . . . . .                                                        | 36           |

|       |                                                                                                                             |            |
|-------|-----------------------------------------------------------------------------------------------------------------------------|------------|
| 3.1.2 | A spline-based GLM captures dLGN spatio-temporal RFs and their modulation by CT feedback and behaviour . . . .              | 40         |
| 3.1.3 | CT feedback is enhanced in the absence of a patterned visual stimulus . . . . .                                             | 46         |
| 3.1.4 | CT feedback affects stimulus decoding . . . . .                                                                             | 49         |
| 3.2   | <b>Study II:</b> Low-firing, bursting patterns in lateral habenula neurons are linked to behavioral modulation. . . . .     | 59         |
| 3.2.1 | LHb neurons spontaneously fire in four distinct patterns .                                                                  | 61         |
| 3.2.2 | Firing pattern Type-1 LHb neurons are preferentially modulated by noxious stimuli . . . . .                                 | 63         |
| 3.2.3 | Type-1 firing pattern LHb neurons are preferentially modulated by spontaneous behavioral state . . . . .                    | 64         |
| 3.3   | <b>Study III:</b> Plasticity of neuronal dynamics in the lateral habenula for cue-punishment associative learning . . . . . | 69         |
| 3.3.1 | Punishment drives distinct LHb neuronal populations with excitation or inhibition . . . . .                                 | 71         |
| 3.3.2 | Disrupting LHb function affects cue discrimination during punishment associative learning . . . . .                         | 72         |
| 3.3.3 | LHb neuronal populations develop opposing responses to punishment-predictive cues across learning . . . . .                 | 75         |
| 3.3.4 | Acetylcholine, but not glutamate, GABA and serotonin, release in LHb mirrors neuronal dynamics after conditioning           | 79         |
| 3.4   | <b>Study IV:</b> Juxtacellular recordings from identified neurons in the mouse locus coeruleus . . . . .                    | 86         |
| 4     | <b>DISCUSSION</b>                                                                                                           | <b>104</b> |
| 4.1   | Summary . . . . .                                                                                                           | 104        |
| 4.2   | The challenge of balancing detail and utility . . . . .                                                                     | 105        |

---

|                                                                  |            |
|------------------------------------------------------------------|------------|
| 4.3 Our brains, our future? . . . . .                            | 109        |
| <b>5 Acknowledgements</b>                                        | <b>112</b> |
| <b>Bibliography</b>                                              | <b>113</b> |
| <b>Appendix A Supplementary material for Study I:</b>            |            |
| <b>Schmors, Kotkat, and Bauer et al. (2023)</b>                  | <b>175</b> |
| A.1 Supplementary Figures . . . . .                              | 175        |
| A.2 Materials and methods . . . . .                              | 183        |
| A.2.1 Surgical procedures . . . . .                              | 183        |
| A.2.2 Extracellular multi-electrode array (MEA) recordings . . . | 185        |
| A.2.3 Locomotion . . . . .                                       | 185        |
| A.2.4 Eye tracking . . . . .                                     | 185        |
| A.2.5 Optogenetic feedback suppression . . . . .                 | 186        |
| A.2.6 Joint modulation by additional inputs . . . . .            | 186        |
| A.2.7 Visual stimulation . . . . .                               | 186        |
| A.2.8 Histology . . . . .                                        | 187        |
| A.2.9 Control experiments with Ntsr-negative mouse . . . . .     | 188        |
| A.2.10 Spike sorting and unit extraction . . . . .               | 188        |
| A.2.11 Data analysis . . . . .                                   | 188        |
| A.2.12 Spline-based generalized linear model . . . . .           | 190        |
| A.2.13 Decoding analysis . . . . .                               | 192        |
| <b>Appendix B Supplementary material for Study II:</b>           |            |
| <b>Zouridis and Schmors et al. (2024)</b>                        | <b>194</b> |
| B.1 Supplementary Figures . . . . .                              | 194        |
| B.2 Materials and methods . . . . .                              | 198        |
| B.2.1 Experimental animals. . . . .                              | 198        |

|       |                                                                                              |     |
|-------|----------------------------------------------------------------------------------------------|-----|
| B.2.2 | In vivo juxtacellular recording and labeling in anesthetized mice. . . . .                   | 198 |
| B.2.3 | In vivo juxtacellular recordings and orofacial videography in awake head-fixed mice. . . . . | 198 |
| B.2.4 | Quantification and statistical analysis. . . . .                                             | 199 |

## **Appendix C Supplementary material for Study III:**

|                             |                                                      |            |
|-----------------------------|------------------------------------------------------|------------|
| <b>Congiu et al. (2023)</b> |                                                      | <b>201</b> |
| C.1                         | Supplementary Figures . . . . .                      | 201        |
| C.2                         | Materials and methods . . . . .                      | 208        |
| C.2.1                       | Experimental subjects . . . . .                      | 208        |
| C.2.2                       | Viruses . . . . .                                    | 208        |
| C.2.3                       | Stereotaxic surgeries for viral injections . . . . . | 208        |
| C.2.4                       | Chronic implants . . . . .                           | 208        |
| C.2.5                       | Histology and immunohistochemistry . . . . .         | 209        |
| C.2.6                       | Pavlovian conditioning . . . . .                     | 209        |
| C.2.7                       | Two-photon microscopy . . . . .                      | 210        |
| C.2.8                       | Fiber photometry recordings . . . . .                | 211        |
| C.2.9                       | In vivo electrophysiological recordings . . . . .    | 211        |
| C.2.10                      | In vitro electrophysiology . . . . .                 | 212        |
| C.2.11                      | Real time place aversion/preference test . . . . .   | 213        |
| C.2.12                      | Quantification and statistical analysis . . . . .    | 213        |

## **Appendix D Supplementary material for Study IV:**

|                                      |            |
|--------------------------------------|------------|
| <b>Zouridis et al. (2024)</b>        | <b>214</b> |
| D.1 Supplementary Figures . . . . .  | 214        |
| D.2 Materials and methods . . . . .  | 221        |
| D.2.1 Experimental animals . . . . . | 221        |

|       |                                                                                          |     |
|-------|------------------------------------------------------------------------------------------|-----|
| D.2.2 | Surgical procedures . . . . .                                                            | 221 |
| D.2.3 | Juxtacellular recording and labelling . . . . .                                          | 221 |
| D.2.4 | Histological analysis, morphological reconstruction and<br>cell quantification . . . . . | 222 |
| D.2.5 | Electrophysiology data analysis . . . . .                                                | 223 |
| D.2.6 | PCA embedding . . . . .                                                                  | 224 |
| D.2.7 | Logistic regression . . . . .                                                            | 225 |
| D.2.8 | Statistical analysis . . . . .                                                           | 225 |

# List of Figures

|     |                                                                                                                                                            |    |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 | Statistical frameworks for sensory processing . . . . .                                                                                                    | 3  |
| 2.1 | Feedforward and feedback connections in the visual system . . . . .                                                                                        | 7  |
| 2.2 | Neural circuitry for reward and aversion . . . . .                                                                                                         | 9  |
| 2.3 | Finding the right level of abstraction . . . . .                                                                                                           | 12 |
| 2.4 | Contextual modulation of sensory processing by behavioral state<br>or arousal . . . . .                                                                    | 15 |
| 2.5 | Schematic overview of clustering algorithms . . . . .                                                                                                      | 30 |
| 3.1 | CT feedback and behaviour modulates dLGN responses to movies                                                                                               | 39 |
| 3.2 | Spline-based GLM captured the RF, the influence of CT feedback<br>and the impact of behaviour on responses of dLGN neurons . . . . .                       | 42 |
| 3.3 | Adding model predictors for modulatory inputs improves the per-<br>formance for a subgroup of poorly visually responsive dLGN neu-<br>rons . . . . .       | 47 |
| 3.4 | The effect of CT feedback is dependent on the presence or absence<br>of the visual stimulus . . . . .                                                      | 50 |
| 3.5 | Two-alternative-forced-choice decoder shows better stimulus dis-<br>crimination for CT feedback modulated neurons when feedback<br>is suppressed . . . . . | 51 |
| 3.6 | GMM clustering on spontaneous firing features reveals four LHb<br>neuron firing pattern types . . . . .                                                    | 62 |
| 3.7 | Sensory and behavioral modulation of LHb neurons with different<br>firing spontaneous pattern types . . . . .                                              | 65 |
| 3.8 | LHb neurons can show excitatory or inhibitory responses to foot<br>shock . . . . .                                                                         | 73 |
| 3.9 | Kinetics heterogeneity of FS-excitatory responses in LHb neurons                                                                                           | 74 |

## List of Figures

---

|      |                                                                                                                             |     |
|------|-----------------------------------------------------------------------------------------------------------------------------|-----|
| 3.10 | LHb integrity is necessary for cue discrimination during Pavlo-<br>vian conditioning . . . . .                              | 75  |
| 3.11 | Single neuron inhibitory and excitatory responses adapt during<br>associative learning . . . . .                            | 78  |
| 3.12 | Presynaptic ACh release in the LHb, but not glutamate, GABA,<br>or serotonin, reflects cue discrimination . . . . .         | 80  |
| 3.13 | In vivo recordings from anatomically identified TH-positive and<br>TH-negative neurons within the mouse LC complex. . . . . | 91  |
| 3.14 | Electrophysiological properties of TH-positive and TH-negative<br>neurons . . . . .                                         | 93  |
| 3.15 | Separating TH-positive and TH-negative cells using sparse PCA<br>and a logistic regression classifier . . . . .             | 95  |
| 3.16 | Foot shock responses of TH-positive and TH-negative neurons . .                                                             | 97  |
| 4.1  | Topology of "NeuroAI" . . . . .                                                                                             | 110 |
| A.1  | Efficiency of optogenetic manipulation of CT feedback . . . . .                                                             | 176 |
| A.2  | CT feedback modulated neurons are closer to each other in the<br>mouse dLGN . . . . .                                       | 176 |
| A.3  | CT feedback and behaviour modulates dLGN responses to movie<br>clips . . . . .                                              | 177 |
| A.4  | Spatial RF estimation by the spline-GLM model . . . . .                                                                     | 178 |
| A.5  | Spline-based GLM captures spatio-temporal RFs of different func-<br>tional cell types in dLGN . . . . .                     | 179 |
| A.6  | Spline-GLM model kernels for modulatory inputs . . . . .                                                                    | 180 |
| A.7  | Model performance depends on various factors . . . . .                                                                      | 181 |
| A.8  | Influences of visual stimulus on effects of CT FB suppression in<br>dLGN . . . . .                                          | 182 |

|     |                                                                                                                    |     |
|-----|--------------------------------------------------------------------------------------------------------------------|-----|
| B.1 | Comparison of LHb neurons recorded under ketamine or urethane anesthesia . . . . .                                 | 195 |
| B.2 | Distributions of features and their correlations across LHb neurons in anesthetized and awake conditions . . . . . | 196 |
| B.3 | Comparing LHb neurons recorded under anesthetized condition and awake condition . . . . .                          | 197 |
| C.1 | Features of the Pavlovian discrimination task . . . . .                                                            | 202 |
| C.2 | Optical inhibition of LHb during the Pavlovian discrimination task                                                 | 203 |
| C.3 | Immunolabeling of glutamatergic and GABAergic neurons in the LHb . . . . .                                         | 204 |
| C.4 | Functional and anatomical signatures of LHb neurons during conditioning . . . . .                                  | 205 |
| C.5 | Potentiation of synaptic excitation in the LHb after learning . . . .                                              | 206 |
| C.6 | Analysis of neurotransmitter dynamics in the LHb . . . . .                                                         | 207 |
| D.1 | Histological analysis of the mouse LC complex . . . . .                                                            | 215 |
| D.2 | Histologically identified neurons within the LC complex . . . . .                                                  | 216 |
| D.3 | Morphological properties of TH-positive and TH-negative neurons                                                    | 217 |
| D.4 | Spike waveforms of the identified neurons in the mouse LC complex                                                  | 218 |
| D.5 | Footshock responses of identified neurons in the mouse LC complex                                                  | 219 |
| D.6 | Colocalization of TH with additional noradrenergic markers (DBH and NET) . . . . .                                 | 220 |

# Notation

## Abbreviations

|                    |                                                 |
|--------------------|-------------------------------------------------|
| <b>AGI</b> .....   | Artificial general intelligence                 |
| <b>AI</b> .....    | Artificial intelligence                         |
| <b>BIC</b> .....   | Bayesian information criterion                  |
| <b>CNN</b> .....   | Convolutional neural network                    |
| <b>CS</b> .....    | Conditioned stimulus                            |
| <b>dLGN</b> .....  | Dorsolateral geniculate nucleus of the thalamus |
| <b>DNN</b> .....   | Deep neural network                             |
| <b>GLM</b> .....   | Generalized linear model                        |
| <b>GMM</b> .....   | Gaussian mixture model                          |
| <b>LC</b> .....    | Locus coeruleus                                 |
| <b>LHb</b> .....   | Lateral habenula                                |
| <b>PCA</b> .....   | Principal component analysis                    |
| <b>RGC</b> .....   | Retinal ganglion cell                           |
| <b>RPE</b> .....   | Reward prediction error                         |
| <b>t-SNE</b> ..... | t-distributed stochastic neighbour embedding    |
| <b>US</b> .....    | Unconditioned stimulus                          |
| <b>VTA</b> .....   | Ventral tegmental area                          |

# 1 | INTRODUCTION

In this thesis, I will discuss how statistical tools and mathematical models can be used to analyze neural activity. Particularly, I am interested in identifying patterns in the data and how that can help to make sense of the underlying neural mechanisms. I will focus on sensory processing, specifically how visual information is represented by neural activity, as well as how it is modulated by behavioral state and plasticity. All data sets analyzed in this thesis were recorded from the mouse brain. Mice share a high degree of genetic similarity with humans and are also available for genetic manipulations. This allows the experimenter to create specific gene knockouts and study their effect on neural activity and behavior. Their basic brain structure and function is similar to ours, especially with respect to sensory processing and learning. All mice studied for the scientific studies contained in this thesis were kept according to the guidelines and regulations of the European Communities Council Directive [86] and the German Law for Protection of Animals [58] to ensure their well-being and ethical treatment. This thesis is a step forward toward using statistical models to simulate living neural systems. However, due to the complexity of the brain and its functions we are still dependent on animal data to refine and validate alternative methods.

## **Motivation**

How we perceive reality shapes everything we do. It determines our actions, how we feel, and what we think. Understanding the underlying neural mechanisms is a pressing goal if we want to understand how this information is represented and processed in the brain. The human brain, in particular, seems to be particularly successful in managing information, learning new concepts, and adapting behavior to ensure the organism's survival. In terms of biomass, humans on this planet outweigh all other wild mammals combined by an order of magnitude [30]. What

## 1. INTRODUCTION

---

makes our brains so successful? To understand the neural dynamics underlying perception and learning, it becomes increasingly important to apply diverse statistical tools and models to the ever-growing data sets collected by experimentalists. Statistical models of neural data are indispensable if we want to generate insights from the data and discover the underlying mechanisms of brain function.

### **Aim and Objectives**

In this thesis, I explore how a diverse set of statistical methods and computational modelling can be used to explore fundamental brain mechanisms, including the processing of sensory information by single neurons and their populations, cell type identity, and the influence of behavior and 'brain states'. While mechanistic and biophysical models, such as the Hodgkin-Huxley model [166], are important and widely used approaches for study neural phenomena, I will focus on statistical and phenomenological models in this thesis and the included studies. These models aim to find patterns and make predictions in order to describe relationships between relevant variables in the system. Particularly, I will discuss how sensory and behavioral inputs are integrated, how these inputs are represented at the cellular level and how they influence animal behavior and neural plasticity. The studies presented in this thesis are the result of close collaborations with experimentalists, who thoroughly recorded data that encompasses the functional, structural, and molecular properties of neurons in the mouse brain. Overall, this thesis aims to advance our knowledge of neural dynamics and its modulation by both internal states and external stimuli using statistical frameworks.

### **Thesis Outline**

I will start by providing background on (1) principles of neural processing ([Figure 1.1 left](#)) and (2) statistical frameworks for neural data ([Figure 1.1 right](#)). Regarding

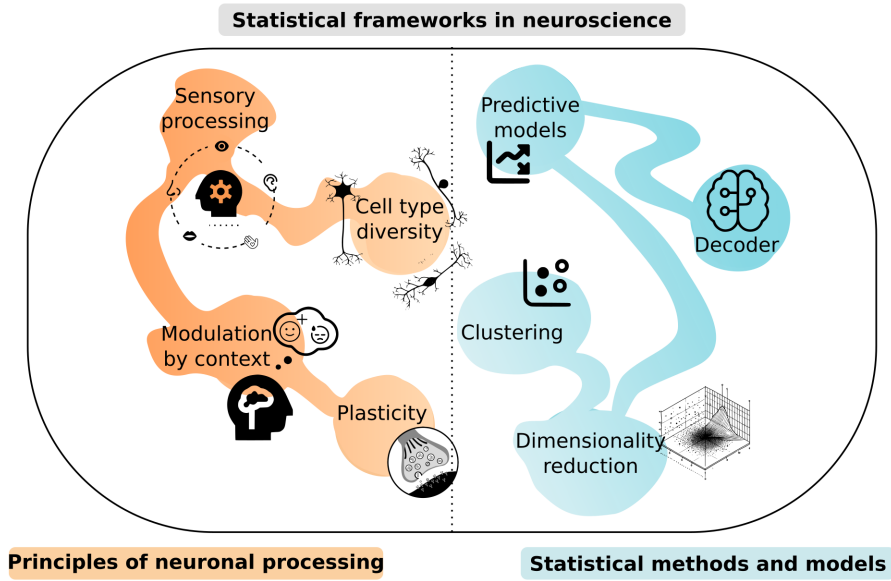


Figure 1.1: **Thesis outline.** Applying statistical frameworks to neuroscience data requires background knowledge in two areas: the biological basis of neural processing (*left, orange*) and statistical methods and models (*right, blue*). Important principles of neural processing include (from *top to bottom*) the perception of our environment ("Sensory processing", section 2.1.1), the computational significance of cell types ("Cell type diversity", section 2.1.2), the contextual modulation of neuronal activity ("Modulation by context", section 2.1.3), and plasticity and associative learning ("Plasticity", section 2.1.4). Statistical methods and models relevant for this thesis are (from *top to bottom*) Linear and non-linear predictors ("Predictive models", section 2.2.1), "Decoders" (section 3.1.4), "Clustering" (section 2.2.3), and "Dimensionality reduction" techniques (section 2.2.4). There are multiple synergies between both fields. Linear predictors can model the relationship between sensory input and neural activity. Decoders can infer stimuli from neural activity. Cell types can be identified by clustering algorithms and clustering can also be based on plasticity profiles. Dimensionality reduction techniques can identify latent factors as sources for neural variability. There are also multiple connections of these aspects within one field: e.g. different cell types process different sensory input differently (*orange connections*), context influences the processing of sensory information, and contextual modulation such as attention can drive the synaptic plasticity in neural circuits. On the other hand (*blue connections*), linear regression can be used for example to predict sensory inputs from neural data, or dimensionality reduction techniques can extract the most relevant features that then function as inputs for predictive models or for clustering algorithms. The aspects in this figure are by far not extensive but represent an overview of the topics relevant for this thesis and the included studies.

## 1. INTRODUCTION

---

the biological background (section 2.1), I will discuss how sensory information is perceived by the brain (section 2.1.1) and how the identity of cell types plays an important role in brain function (section 2.1.2). I will further describe how neural activity is modulated by the external and internal context (section 2.1.3) and how neural circuits can be further adapted by plasticity and associative learning (section 2.1.4). With respect to the statistics background (section 2.2), I will explain how linear predictors and decoders can be used to either predict neural activity (section 2.2.1) or inform about the stimulus given neural activity (section 2.2.2). I will further describe how clustering can be used to identify neuronal cell types (section 2.2.3) and how methods of dimensionality reduction methods can be helpful as a tool to extract features from neural data or to visualize a data set (section 2.2.4).

While detailed descriptions of the studies that constitute this thesis will follow in section 3.1, section 3.2, section 3.3, and section 3.4, I outline a list of publications below:

**Study I** — investigate the effects of non-retinal inputs on the visual responsiveness of neurons in the dorsolateral geniculate nucleus of the thalamus

**Study II** — characterizes the electrophysiological diversity of neurons in the lateral habenula and their unique responses to noxious stimuli and behavioral states

**Study III** — investigates the neural mechanisms underlying threat association in the lateral habenula

**Study IV** — shows that noradrenergic and non-noradrenergic neurons within the locus coeruleus can be classified by their distinct electrophysiological properties

## 2 | BACKGROUND

### 2.1 Principles of neural processing

This section will provide a broad overview of the principles of neural processing in the mammalian brain. Often the sections contain examples from the two brain areas discussed in this thesis: the dorsolateral geniculate nucleus (dLGN) and the lateral habenula (LHb). By "Principles of sensory processing", I refer to the fundamental mechanisms and concepts important in the context of this thesis by which sensory systems receive, transform, and interpret external stimuli into meaningful information that the brain can use to understand and interact with the environment.

#### 2.1.1 Sensory processing

Animals and humans rely on sensory input to interact with the world, to plan and to survive. The processing of sensory information (such as visual, auditory, tactile, olfactory, and gustatory systems) refers to the way the nervous system receives, interprets, and responds to sensory information from the environment. This involves the detection of sensory stimuli through specialized receptors, its conversion into neural signals, and the subsequent integration by the brain. In this section, I will focus on the processing of visual and aversive (i.e. negative) stimuli as they are the most important sensory inputs for Study I-IV.

#### **Vision: feedforward and feedback processing**

Visual processing is a hierarchical process that involves both, feedforward and feedback connections in the brain. For the forward pathway, visual information is initially perceived by the retina, transmitted to the thalamus, and eventually

## 2. BACKGROUND

---

reaches the in the visual cortex. From here, the information is send to higher brain areas involved in decision making. The feedback pathways are divers and include corticothalamic feedback [54], cortico-cortical feedback [15], and intra-areal feedback [47]. The dorsolateral geniculate nucleus of the thalamus (dLGN), a central hub for visual processing, plays a multifunctional role by combining inputs from multiple visual and non-visual areas [433]. Regarding visual inputs, the layered organization of the dLGN allows for a parallelized processing of visual features such as color, contrast, or motion since each layer receives inputs from different types of retinal ganglion cells [reviewed in 100, 101]. This topographic organization is also the basis for the receptive fields in dLGN, which is the area (i.e. "field") and a particular visual feature that the neuron is most responsive to [111]. The activity of neurons in dLGN is also modulated by arousal, which changes the transmission of visual information depending on the context and the physiological state [262, 271, 281, 378, 433, see also section 2.1.3]. Despite the influence of behavior on responses of dLGN neurons to visual input, it is the corticothalamic feedback that presents the major source modulation of activity in dLGN [292]. It is a top-down modulation of feedforward visual information, which is suggested to play a role in the refinement of the spatial and temporal properties of receptive fields [164, 369, review], the integration of contextual information [15], and attention [55, 257]. The feedback connections between cortex and dLGN involve a reciprocal interaction, where cortical activity influences thalamic processing, and thalamic activity, in turn, influences cortical processing ([333] and see **Figure 2.1**).

In the past, the dLGN was viewed as a simple relay station for visual information [363], however this view is recently debated and will be further discussed in Study I (see section 3.1).

The thalamus is also a gateway for sensory signals, including aversive cues, to

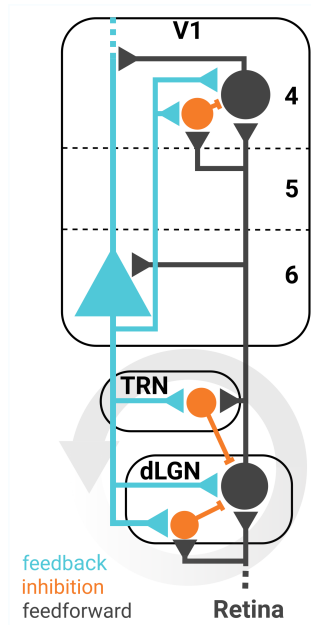


Figure 2.1: **Feedforward and feedback pathways in the visual system.** Light information is first received by the retina (*bottom*) and then send via the dLGN to layer 4 of the primary visual cortex (V1; *top*). Layer 4 neurons project onto neurons with their soma in layer 6 (*blue*), which send direct and indirect (through the thalamic reticular nucleus; TRN) feedback connections to dLGN. Feedforward signal in *gray*, feedback excitatory connections in *blue*, and feedback excitatory connections in *orange*. This figure was modified from Sillito (2006) [369].

higher brain areas involved in processing of negative emotions [461] as discussed in the next section.

### **Aversion: processing of negative percepts**

Learning from mistakes is a critical survival skill for all animals and humans. Processing of aversive experiences is mediated by the reward prediction error (RPE), that can be either code for a reward that exceeds expectations (positive RPE) or for less reward than expected (negative RPE). Encoding of aversive behavior requires a coordination of many brain areas including the amygdala, the lateral habenula (LHb), and the ventral tegmental area (VTA). While a sophisticated coordination of all involved brain nuclei is necessary, this paragraph will focus on the functional role of the LHb within the scope of aversive processing.

## 2. BACKGROUND

---

The LHb is known as the 'anti-reward' system in the brain [172]. While the amygdala is the hub for negative emotions and fear responses [216], the LHb is associated with coding for negative valence and activity refinement and modulation in neuronal reward circuits. The LHb connects the midbrain with forebrain and is strategically positioned to control both, the serotonergic and dopaminergic systems<sup>1</sup> [76, 156, 180, 181, 255, 383, 388, 425]. Activity in the LHb increases in response to unexpected negative stimuli [413] or the absence of expected positive stimuli [335] (in human studies this can be simple stimuli such as happy or sad smiley faces). However, neurons respond to negative stimuli in heterogeneous ways: Congiu et al. [83] found 78% of the neurons respond with an increase, the others with a decrease in activity. This heterogeneity will be further explored by Study II (section 3.2).

The LHb has reciprocal connections with the VTA, which means that it negatively regulates the activity in the VTA and is itself regulated by the VTA through bidirectional connections (see **Figure 2.2**). The VTA is the major source of dopamine production in the brain and activated by rewarding stimuli including food, sex and other pleasurable experiences (positive RPE) [347]. Aversive processing can have impact on the overall perceptual experience by mediating the learning process through negative or positive valence to discriminate between visual [350] or olfactory [227] sensory cues that were indistinguishable before. The multimodal nature of aversive processing is also apparent in the input connections that the LHb receives from sensory pathways such as the thalamus, transmitting cues of aversive visual stimuli [426]. It also receives suppressing input from ventral LGN [173], part of non-image-forming visual pathway, that may mediate the antidepressant

---

<sup>1</sup>The serotonergic and dopaminergic system in the brain refer to a network of neurons that use serotonin or dopamine as their primary neurotransmitter. The major source of dopamine in the brain is the VTA. Both systems are involved in mood regulation, motor control, reward processing, and cognition [37, 251].

effects of bright light suppressing LHB activity. Through its pivotal position in the brain, the LHB seems to be involved in many motivational, movement, and cognitive functions [285], but also sleep [29], pain [358], and navigation [160].

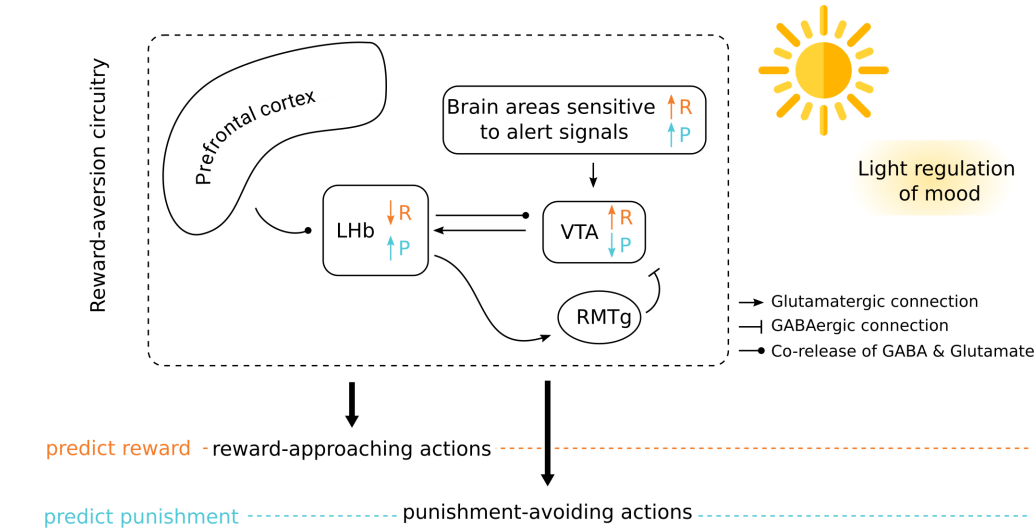


Figure 2.2: **Neural circuitry for reward and aversion.** The circuitry processing aversive stimuli involve complex connections of multiple brain areas, among them the the lateral habenula (LHb), the rostromedial tegmental nucleus (RMTg), and the ventral tegmental area (VTA). Negative experiences (punishment, P; *blue*) increases the activity in LHb, while positive stimuli (reward, R; *orange*) decrease activity in LHb (indicated by arrows). The opposite is true for the VTA, the major source of dopamine in the brain. The LHb receives input from the forebrain and suppresses activity in the VTA through indirect connections via RMTg. LHb and VTA are further connected through reciprocal projections which can have both, suppressing (inhibitory) and activating (excitatory) effects. Inhibitory effect are implemented through Glutamatergic connections and excitatory effects are implemented through Glutamatergic connections. The LHb also receives input from light sensitive areas such as the retina and the visual thalamus.

For a regulation of all these functions, a sophisticated balance of activation and suppression of and by other brain areas involved in the processing of negative stimuli is critical. A dysregulation of the reward-aversion circuitry can lead to serious mental health conditions [119, 288, 321, 448], such as mood disorders, schizophrenia, and substance use disorder. For example in major depression, depressive-states are mainly associated with an hyperactivity in LHb [64, 274]. Artificially enhancing or suppressing activity in the LHb can induce depressive or anti-depressive effects [90, 214, 226, 447]. These observations were made for multiple species, including humans, where pilot studies reported full remis-

## 2. BACKGROUND

---

sion of depression in patients using deep brain stimulation to inactivate the LHb [342]. Consequently, the LHb is the main molecular target of anti-depressants [91, 233, 447]. Emerging directions to leverage drug research should thus include discovering the relationships between the LHb and other brain regions involved in regulating negative emotions as well as the characteristic firing patterns of LHb neurons.

In this thesis, I present new research that discovers more cellular diversity in the LHb in response to negative stimuli, as well as the adaptive nature of neural responses through conditioning (see Study II in section 3.2). Further, the role of neurotransmitters at single-cell level in the LHb during associative learning will be further discussed in Study III (see section 3.3).

### 2.1.2 The computational significance of cell type diversity

Classification of neurons into cell types is necessary for discovering and understanding mechanisms in the brain [48, 124, 353, 384]. Cell type classification also allows us to identify cells based on their genetic factors and to make them accessible for modification (see Study I in section 3.1 for optogenetic approach). It also makes it possible to look at differences between species and gain insights into evolution [16] and identify cell-type specificity of diseases and possibly precisely target them. Further, it is a form of dimensionality reduction because it allows us to look at groups of neurons that share features instead of looking at single neurons. Diverse cell types also have a significant role in neuronal computation since the computational capacity of the brain is determined by its cellular composition [439]. From a computational theory perspective, the differentiation into cell types provides for information processing with optimal capacity and energy efficiency using parallel but specialized processing pathways [250]. The development of neurons with different functions is an evolutionary necessity in response

to complex and changing environments. This is analogous to what drives the development of communication and information systems and hardware such as GPUs but also artificial neural networks. Similar to parallel computing in programming and hardware, the processing of information through parallel feature channels allows for a fast and efficient perception with different neuron types responsible for specific tasks in the network [77]. It allows the brain to be adaptive and salable to changing environments and learning experiences, which is a key feature of successful biological and artificial systems [245].

Often, the function of a single neuron only emerges in the context of the circuit it is embedded in and it is hard to define the function of a type. However, there are three main criteria<sup>2</sup> based on which neurons can be grouped into populations [455]:

1. functional properties (physiology)
2. structural properties (morphology)
3. molecular properties (genetic profile)

Functional properties of neurons can be for example the resting potential or firing rate. Structural properties can be the soma size or branching patterns, and the genes determine the protein composition of neurons.

But even using these criteria to group neurons into types, no cell type is homogeneous. There are always variations even within the same defined cell type. Even if two cells have the same gene expression profile, they will diversify due to environmental and stochastic processes [14, 43]. Physiological properties of neurons can for example vary with temperature or the number of synaptic inputs

---

<sup>2</sup>Using connectomics as a fourth criterion is undoubtedly an important and promising direction. However the data was not available for the studies discussed below and this criterion will thus not be discussed in this thesis.

## 2. BACKGROUND

---

even if the molecular barcode of two neurons stays the same. Hence, cell types are also a dynamic concept due to transcriptomic alterations [98] or neurotransmitter switching [379]. On the other hand, even for more permanent features such as genetic factors, not all genes that are present also need to become actually expressed. Continuous versus discontinuous distributions of cell types become visible using data visualization methods (see section 2.2.3) and will be further discussed in 4.2 Discussion.

One could argue that a sophisticated approach of cell type characterization should be based on all three criteria combined (functional, structural, and molecular). However the prerequisite for this would be that all of them co-vary. While this has been shown for some cases, it is not true for all and can make an adequate cell type characterization challenging. Further, this would require collecting multiple datasets for the same neuron which is very challenging.

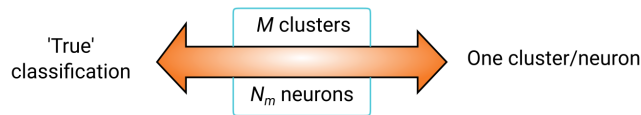


Figure 2.3: **Finding the right level of abstraction.** Neuronal classification has the challenge of finding the balance between either splitting  $N$  neurons into too many groups ( $M$  too large) or lumping cells into too broad groups ( $M$  too small). Extreme splitting would eventually lead to each neuron comprising its own cluster (*right*). Indeed no two neurons are exactly identical. However, the goal should be to find the optimal number of clusters  $M$  and get as close as possible to the ground truth ("True" classification, *left*). It is questionable if the optimal number of clusters even exists because neurons are often described by a gradient of diverging feature properties. The figure was modified from Armañanzas et al. (2015), Fig. 3 [17].

In the visual system, sensory processing starts in the retina, where the optic nerve consists of retinal ganglion cells (RGC) exhibiting parallel feature channels [26]. The different types of RGCs cover a broad range of visual features: Specific RGC types respond, for example, to movements in different directions or prefer different light levels. Another example is a group of RGCs that is suppressed by strong contrasts. Some RGC types can be linked to genetic factors, while others show

distinct characteristics in their morphological characteristics (e.g. cell body size or dendritic stratification). This shows that having diverse cell types leads to processing of information in the visual system not in a pixel-wise manner like in a camera but rather resembles the processing of ANNs such as ImageNet [205].

Why has the brain evolved to use diverse neuronal cell types and processing strategies? One way to look at this is to compare how biological and artificial systems solve the same task such as object recognition. DiCarlo et al. (2012) [104] argue that the algorithm for object recognition is based on hierarchical processing of abstraction layers (retina, thalamus, visual cortex) but with specialized neuronal types at each stage. This is essential for the ability to extract and integrate visual information and especially for experience-dependent plasticity in differentiated neuronal cell types. It allows the recognition system to achieve invariance to changes in position and scale of the object, while still being able to generalize across different instances of objects. Again, this could not be achieved without specialized processing pathways, implemented by diverse neuronal cell types.

In the future, progress in characterizing neuron cell types will depend on technical advances to better quantify the molecular and physiological properties simultaneously of large groups of neurons, and especially the connectivity between neurons. This is necessary to ultimately understand the complex functions of single neurons and their identity in the circuit.

### **2.1.3 Contextual modulation of neural activity**

Responses of neurons depend heavily on the context. Thus, the perception of sensory information is not a passive process but is always influenced by context (e.g. attention, motivation, and expectation) and the behavioral state (e.g. active versus quiescent). This interplay between sensory processing and the animal's

## 2. BACKGROUND

---

state influences its perception and actions.

Requirements for neural resources drastically change from sleep to wakefulness and from quiet to active states. In rodents, a high arousal state coincides with a larger pupil size and more active locomotion and influences many sensory areas of the brain e.g. the somatosensory cortex, the auditory cortex, or the visual cortex (for review see [60]). In the visual system, arousal was shown to support robust encoding [35] and enhance spiking in the visual cortex [287]. Particularly influenced are neurons with high frequency receptive fields [268], which would suggest that it helps to increase the spatial resolution during active states [268]. Activity-dependent modulations play a role not only in the cortex but also along the processing pipeline as early as the thalamus [116]. The dLGN, as part of the thalamus, has a central position in the brain and has many reciprocal connections (see section 2.1.1). Traditionally, it was viewed as a gatekeeper and modulator of sensory information before it reaches the cortex [79]. Erisken and colleagues showed that the dLGN in mice carries locomotion signals and consequently seems to encode a combination of visual and behavioral information in its neural activity [116]. Further, the thalamus was also suggested to mediate functional connections within and across the cortex, potentially controlling the “computational-state” of the cortex (Figure 2.4). Tuning the cortical activity in this context-dependent way would add cognitive flexibility and enhance the encoding of relevant sensory information [279]. Activity-dependent modulations are also likely mediated through feedback projections providing top-down context information that are governed by attention (see also section 2.1.1).

The state of activity of an animal can be measured from high to low based on the activation level of the central nervous system. High levels of arousal are, among other things, associated with strong activation of the cortical circuits through noreadrenalin mediated by the locus coeruleus (LC) and through acetylcholine from the

basal forebrain (reviewed in [219, 262]). They also act directly on the thalamus and could lower the local activity by for example suppressing potassium channels [364] and reducing thalamic bursting defined as fast successions of action potentials [168].

The LC is the primary source for noradrenalin, which is a key neuromodulator that affects the excitability of neurons, synaptic plasticity, and overall neural network dynamics [36]. Neuromodulators act on top of hard wired synaptic connections and adjusting the processing of information in neural circuits to optimize behavior according to environmental demands. Noradrenaline effectively increases the signal-to-noise ratio in thalamic responses, making it easier for the thalamus to transmit salient sensory information during states of high arousal [162]. Taken together, arousal centers in the brain, among them the LC, mediate context-dependent modulation through neuromodulators and influence how information is processed along the sensory pathways during active behavior.

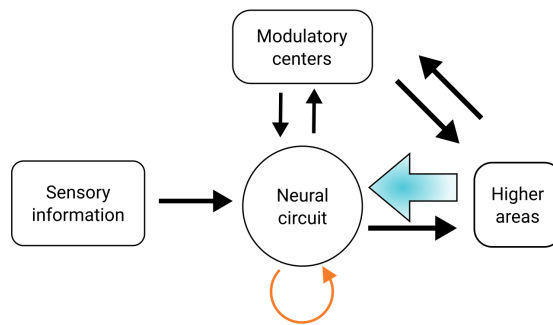


Figure 2.4: **Contextual modulation of sensory processing by behavioral state or arousal.** Schematic representation of synaptical interactions involved in contextual modulation and context-dependent sensory processing (*left*). Higher areas interact with modulatory centers (*top*) and jointly influence activity in neural circuits by top-down modulation (*blue* and black arrows). Further, neural circuits of sensory processing are also influenced by lateral connections within sensory areas (*orange* recurrent arrow). The figure was modified from Batista-Brito et al. (2018), Fig1 [32].

### 2.1.4 Plasticity and associative learning

Associative learning is a form of classical conditioning where an organism learns to associate a previously neutral sensory stimulus (cue) with an valuable outcome (reward or punishment) [217]. This process involves temporally pairing the neutral stimulus (conditioned stimulus, CS) with the reward or punishment (unconditioned stimulus, US). Over time, the organism learns to anticipate the punishment or reward in response to the CS. This form of learning and adaptive behavior is necessary for survival and allows the organism to predict potentially harmful situations.

Neuroplasticity is the underlying neural mechanisms for associative learning that allows the brain to reorganize circuits by adjusting the strength of synaptic connections between neurons, pruning existing synapses, or forming new ones [201]. It allows the adaptation of the response to environmental stimuli, which is crucial for learning and memory [80]. There are multiple brain regions involved in cue punishment learning: the amygdala [217], the hippocampus [247], the prefrontal Cortex [375], the Lateral Habenula (LHb) [255], the locus coeruleus (LC) [338], and the bed nucleus of the stria terminalis [95]. While many studies investigated the role of the amygdala and the hippocampus as a central hub for emotional processing and contextual learning, respectively, I will focus on the LHb and the LC in this dissertation (see also Study II in section 3.2, Study III in section 3.3, and Study IV in section 3.4).

As discussed in section 2.1.1, the LHb is responsible for processing aversive stimuli, representing negative reward prediction errors but also for the modulation of reward circuits [172]. It modulates the dopaminergic system and the serotonergic system, the two systems in the brain responsible for reward processing and mood regulation during punishment learning [255]. Specifically, the LHb gets activated

by stimuli with negative valence and inhibits the VTA, which in turn suppresses the dopamine release (see section 2.1.1). At the same time, activation of the LHb can lead to serotonin release which signals negative events [458]. During associative learning the LHb responds to both the US and the CS and over time starts to respond more strongly to the CS anticipating a negative outcome, which in turn initiates avoidance behavior [409, 424].

The LC on the other hand, is the primary source of noradrenaline in the brain [20, 336]. Noradrenaline facilitates synaptic plasticity in multiple brain regions including the Amygdala and the Hippocampus, which is necessary to form associations between stimuli [148, 338]. The LC receives direct input from the LHb, which allows the LHb to directly influence LC activity and NE release [7]. It also contributes to rapidly updating stimulus-outcome associations during associative learning [401].

Taken together, associative learning, depends on neuroplasticity and reward prediction errors and involves multiple brain regions with the LHb and the LC as central hubs for modulating reward circuits and processing aversive stimuli.

## 2.2 Statistical frameworks for neural data

Statistical models are often used to replicate experimental results and predict experimental results and provide a theoretical framework for the underlying mechanisms. They are frequently applied in sensory systems neuroscience because the experimental input (the stimulus) can be tightly controlled by the experimenter, while the output (the neural activity) is measured in response to the stimulus [442]. Combining these data with knowledge about the system is used to create statistical models that have the aim of optimally describing how neurons process sensory information. This approach is known as systems identification [248]. It often goes

## 2. BACKGROUND

---

under different names such as receptive field estimation [399], spike-triggered analysis [332, 407], reverse correlation [11, 186], or white-noise analysis [75]. These statistical models are widely applied to describe sensory processing in, e.g. the visual system [68, 96] and the auditory system [105, 258]. Usually, the experiments that are carried out to collect data for systems identification use simple and synthetic stimuli, however it is unclear how they translate to a more natural environment that is usually characterized by a more complex statistical structure [370]. In the past years there have been efforts to move towards a more naturalistic setup, where the animal is awake and behaving while less artificial stimuli are used.

### 2.2.1 Linear and non-linear predictors

Any linear model assumes that the relationship between the input and the output of the function  $y = f_{\beta}(x)$  is linear or at least a reasonable approximation. Linear predictors are simple but often form the basis of many statistical methods for describing how the input  $x$  influences the output  $y$  based on the model parameters  $\beta$  [153]. One example from computational neuroscience could be e.g. a light stimulus governing the activity of neurons in the visual system [96]. The input  $x$  in this case is a vector representing the stimulus, which may include some points in time. The output  $y$  is a scalar representing neural response, which might be measured as spike counts [186], membrane potential [56], or instantaneous firing rate [399]. The input  $x$  can be quantitative, such as the light intensity of the visual stimulus or a transformation of that (e.g. log-transform or basis extension). It can also represent combinations of variables. The observed data consists of  $N$  stimulus-response pairs  $(x_i, y_i)$ , where  $i = 1, 2, \dots, N$ . Using the linear model, we can make predictions about neural activity:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (2.1)$$

Here,  $\hat{y}$  represents the predicted response,  $\beta_0$  is the intercept, and  $\beta_1, \beta_2, \dots, \beta_n$  are the parameters.  $x_1, x_2, \dots, x_n$  are the independent variables.

Typically, we use the training observations  $(x_i, y_i)$  to estimate the model parameters  $\beta$ . The most commonly used estimation method is Ordinary Least Squares, which minimizes the sum of the squared differences (residuals) between the observed and the predicted value:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (2.2)$$

where  $X$  is the matrix of input data with each row representing an observation and each column representing a variable. This formula calculates the best-fitting line by minimizing the residual sum of squares (RSS), minimizing the error in predictions:

$$\min RSS(\beta) = \|y - X\beta\|^2 \quad (2.3)$$

We can also define parameter estimation as an optimization problem. Here, we try to maximize the likelihood of the observed data under the model:

$$\beta^* = \arg \min_{\beta} \left( - \sum_{i=1}^N \log p(y_i | f_{\beta}(x_i)) - \log p(\beta) \right) \quad (2.4)$$

with  $\beta^*$  as the most likely values for the model parameters. This equation effectively captures the essence of a Bayesian approach to parameter estimation, where  $\beta^*$  representing the most likely values for model parameters. The estimation method makes no assumption about the validity of the model, it just finds the

## 2. BACKGROUND

---

best linear approximation of the data.

To discourage overly complex models and enhance generalization, regularization is a crucial technique in model fitting. The two most common forms of regularizations are L1 (lasso) and L2 (ridge) regularisation [153]. L2 regularisation shrinks coefficients towards zero by adding the sum of squared coefficients to the objective function [167] and L1 regularisation sets some of the coefficients to zero by adding the sum of absolute values of coefficients to the objective function [402]. This leads to a sparse model representation and is often used for feature selection.

### Generalized linear models

Because neural responses are often not normally distributed (e.g. firing rates are non-negative), generalized linear models (GLMs) are more appropriate.

GLMs represent an extension to linear models that allows for non-linear relationships between input and output variables [280]. They have a high flexibility in handling data with different distributions, which is important in neural data analysis. The general form is given by:

$$g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \quad (2.5)$$

where  $g(\cdot)$  is the link function,  $\mu_i$  is the expected value of the response variable,  $\beta_0, \beta_1, \cdots, \beta_p$  are the coefficients, and  $x_{i1}, x_{i2}, \cdots, x_{ip}$  are the predictor variables. For normally distributed data the identity link function is used. This is the simplest form of the GLM, which corresponds to a linear model. Other forms use the log link function, which is used for count data that follows a Poisson distribution. This is the case for spike count data in neuroscience [1]. The log link ensures that the predicted values are non-negative:

$$g(\lambda_i) = \log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \quad (2.6)$$

Solving for  $\lambda_i$ , we get:

$$\lambda_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) \quad (2.7)$$

The likelihood of observing a spike is given by:

$$P(Y = y|\lambda) = \frac{\lambda^y e^{-\lambda}}{y!} \quad (2.8)$$

To fit this model, we would typically use maximum likelihood estimation [107]. This involves finding the parameter values that maximize the likelihood of observing the data. The log-likelihood for the Poisson distribution is given by:

$$\ell(\beta_0, \beta_1, \beta_2) = \sum_{i=1}^n (y_i \log(\mu_i) - \mu_i - \log(y_i!)) \quad (2.9)$$

One advantage of GLMs is the interpretability of the coefficients. They indicate the importance of every predictor variable, often called feature. GLMs can also be fit to a succession of time windows and give information about the dynamics of feature importance over time (see Study I in section 3.1).

### 2.2.2 Decoders

The previous chapter (section 2.2.1) focused on predicting neural activity in response to identified stimuli. In this approach, the underlying mechanisms are modeled in order to understand how sensory information is translated (or "encoded") into neural activity. Neurons in the brain, however, are presented with

## 2. BACKGROUND

---

the opposite problem: higher brain areas try to make sense of the information in neural activity presented to them. In other words, they try to "decode" the information from patterns of action potentials. Using statistical approaches to decode the information encoded in neural activity can help to understand (1) the limits of decoding given the variability of neural responses, and (2) to estimate the information content in neural activity. Further, it can be used to compare the precision of information encoding during different conditions, e.g. during different behavioral states (see section 2.1.3).

A neuron will never respond exactly the same even if exactly the same stimulus is presented twice. This inherent stochasticity makes encoding and decoding a problem that can be modelled using the Bayes' theorem [96]. Let  $r$  be the recorded neural response or spike train in response to stimulus  $s$ . The probability of recording the response to a particular stimulus is given by  $P(r|s)$ . This encoding step is usually associated with the statistical model in the system identification approach. Here, the input is the stimulus and the output is the predicted neural response. In the decoding step, on the other hand, we look at the probability of the stimulus being shown given the neural response  $P(s|r)$ . Here, the input is the neural response and the inferred stimulus that was most likely to be shown is the output:

$$P(s|r) = \frac{P(r|s) P(s)}{P(r)} \quad (2.10)$$

with  $P(r)$  as the probability of the response being recorded (assuming that  $P(r) \neq 0$ ) and  $P(s)$  as the probability of the stimulus. Consequently, decoding is only applicable if we know the stimulus statistics during the experiment or in the natural environment. If we assume continuous variables (e.g. an averaged firing rate instead of the discrete number of action potentials) the probabilities are replaced by the probability densities  $p(s|r)$ ,  $p(s)$ , and  $p(r)$ . The neural activity can be the

response of a single neuron or a population of neurons.

After computing the probabilities for all stimuli given the responses, reconstructing the stimulus is then performed by using the maximum or the mean of  $P(s|r)$ .

Multiple other types of decoders have been developed for computational neuroscience, each with a particular application:

1. **Linear decoders** [175, 427]: assume a linear relationship between the predictor and neural activity
2. **Dimensionality reduction techniques** [92, 130]: project neural data into a lower dimensional space, making it easier to decode the most important patterns of the neural activity
3. **Maximum likelihood decoders** [179, 310]: estimate the parameters of a model that maximize the likelihood of the observed neural activity given the stimuli or behavioural outcomes
4. **Artificial neural networks** [31, 135]: capture complex, nonlinear relationships between predictor variables and neural response

This list is not exhaustive and new methods to decode stimulus information from neural responses are continuously being developed. Recently, there has been a trend to focus on neural populations instead of single neurons [23, 299]. This is especially beneficial for robust information encoding because single neurons exhibit strong variability and often have a low signal-to-noise ratio [118, 356]. Additionally, a single neuron receives inputs from multiple converging primary neurons, making population coding the more likely and effective principle for accurately representing information [317, 452].

### 2.2.3 Clustering

In neuroscience, it is often the goal to identify groups ("clusters") of neurons based on similar properties [196]. Clustering is a data-driven unsupervised approach to group observations together in such a way that the pairwise dissimilarities between points of the same group are smaller than those of different groups [153]. This is used for example to group neurons based on their gene expression profiles [343], identify groups based on their responses to a stimulus or a behavioral latent variable [82], or to assign the many action potentials measured during an electrophysiological recording to its respective source neuron (a method called "spike sorting" [322]). Applying clustering algorithms to neuroscience data [17] involves multiple steps:

1. Feature extraction: The data is often multidimensional that needs to be quantified either automatically using dimensionality reduction techniques like principal component analysis (PCA) or hard coded analysis such as e.g. measuring the width of an action potential
2. Algorithm design and similarity metric: To group data points, their similarity in feature space has to be quantified. This might not be straightforward as one has to e.g. deal with temporal correlations when working with time series data.
3. Assigning cluster labels to the identified groups.
4. Validation of clusters: Evaluate if the clusters are meaningful and how many clusters are ideal.

In many cases, it is not only the goal to assign neurons to groups based on quantitative measurements but also to identify the groups themselves that best explain the data [17]. As an example, visual neurons can be typically grouped into ON

and OFF neurons based on their responses to an on-off light stimulus (see 3.1). However, some neurons may respond to both conditions, potentially forming a distinct cluster (ON-OFF neurons; see section 2.1.2).

Common clustering algorithms applied in neuroscience include Gaussian mixture models, K-means, and Hierarchical clustering. They belong to combinatorial algorithms [153] and have the number of clusters  $K$  with  $k \in \{1, \dots, K\}$  prespecified. The goal is to assign each observation  $i \in 1, \dots, N$  to one of the clusters  $k = C(i)$  by minimizing a loss function.

### K-means clustering

K-means clustering is an algorithm that returns a cluster assignment  $C$  and a set of cluster means (centroids)  $\{m_1, \dots, m_K\}$  [153]. In order to find the optimal  $C^*$ , the dissimilarity between each observation and the cluster center is measured using the squared Euclidean distance as the dissimilarity metric.

$$d(x_i, m_K) = \|x_i - m_K\|^2 \quad (2.11)$$

It can also be interpreted as the variance within a cluster. The optimal  $C^*$  is found based on minimizing the sum of squared distances (loss function)

$$\min_{C, \{m_K\}_1^K} \sum_{k=1}^K N_k \sum_{C(i)=k} \|x_i - m_k\|^2 \quad (2.12)$$

The optimization problem in K-means is solved by iteratively running the following two steps:

1. Identify the subset of data points which are closer to the respective center than to any other center.

## 2. BACKGROUND

---

2. Compute the mean for the new subset of observations.

$$C(i) = \arg \min_k ||x_i - m_k||^2 \quad (2.13)$$

The initial means are usually set to random and the centroids are then moved around in an iterative setting ([Figure 2.5left](#)). The steps are repeated until convergence.

### Gaussian mixture models

Gaussian mixture clustering ([Figure 2.5middle](#)) is a probabilistic approach and assumes that each data point  $x_i$  was drawn from a mixture of several Gaussian distributions [153]. The entire model is a blend of these Gaussian distributions. The number of Gaussian distributions is denoted by  $K$  and each one is a density described with a centroid  $\mu_k$  and a covariance matrix  $\Sigma_k$ . Each cluster also has a prior probability (or mixing coefficient)  $\pi_k$  that reflects the expected fraction of the dataset that belongs to each cluster. These are the parameters  $\theta$  of the Gaussian mixture model (GMM):

$$\begin{aligned} \theta &= (\pi_k, \mu_k, \Sigma_k) \\ p(x_i|\theta) &= \sum_{k=1}^K \pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k) \end{aligned} \quad (2.14)$$

Similar to K-means clustering, the parameters for the Gaussian mixture model are estimated in an iterative fashion. The algorithm used for updating the parameters is called expectation maximization (EM) algorithm. Usually, the means are initialized randomly and the variances can be set equal to the overall sample variance.

1. Expectation step: The current  $\theta$  are used to give each observation a proba-

bility of belonging to each cluster  $\gamma(z_{ik})$  (sometimes also called weight or responsibility).

$$\gamma(z_{ik}) = \frac{\pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(x_i | \mu_j, \Sigma_j)} \quad (2.15)$$

Observations very close to the center will get probability close to 1 for this cluster and a probability close to 0 for all other clusters. This is considered to be a *soft* assignment compared to the *hard* assignment in K-means clustering.

2. Maximization step: The new cluster centers, covariance matrices, and mixing coefficients are computed. Each data point contributes to the new parameters with its respective weight  $\gamma(z_{ik})$ .

To evaluate how well the GMM explains the whole data, we can compute the likelihood of each observation under the model and then take the product across all data points. Consequently, we find the optimal cluster assignments by maximizing the log-likelihood function:

$$\log p(X | \pi, \mu, \Sigma) = \sum_{i=1}^N \log \left( \sum_{k=1}^K \pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k) \right) \quad (2.16)$$

The iterations are repeated until the changes in the log-likelihood in successive iterations fall under a predetermined threshold.

### **Hierarchical clustering**

In comparison to K-means clustering and GMMs, hierarchical clustering does not depend on the initialization of centers and the numbers of clusters. Rather, all possible numbers of clusters exist in parallel and observations are grouped based on a pairwise dissimilarity measure  $d(x_i, x_j)$ . Common measures include Euclidean distance or Cosine similarity. Hierarchical clustering results in a representation

## 2. BACKGROUND

---

of the dataset where at each level a group is formed based on merging the most similar clusters from the next lower level. At the lowest level each observation forms its own cluster (singleton cluster) and at the highest level all observations fall into the same cluster. There are two main algorithms of hierarchical clustering which are normally distinguished: *agglomerative* and *divisive* clustering (**Figure 2.5right**). Agglomerative clustering is a bottom-up approach where we start at the lowest level and successively merge similar groups together. In contrast, divisive clustering is a top-down approach where we start from the highest level and successively split the cluster with the largest dissimilarity into two clusters. This results in a hierarchy of clusters nested into clusters. In order to decide which two clusters  $U$  and  $V$  to merge, we need a rule that makes these decisions. This is called the linkage criterion and common choices are:

1. Single Linkage: Clusters with smallest distance are merged.

$$d(U, V) = \min(d(u, v) : u \in U, v \in V) \quad (2.17)$$

2. Complete Linkage: Considers the maximum distance between points.

$$d(U, V) = \max(d(u, v) : u \in U, v \in V) \quad (2.18)$$

3. Average Linkage: Uses the average distance to merge clusters.

$$d(U, V) = \frac{1}{|U||V|} \sum_{u \in U} \sum_{v \in V} d(u, v) \quad (2.19)$$

4. Ward's Method: Minimizes the variance within the cluster.

$$d(U, V) = \sqrt{\frac{|U||V|}{|U| + |V|} \sum_{i=1}^n (c_U^i - c_V^i)^2} \quad (2.20)$$

where  $c_U^i$  and  $c_V^i$  are the centers of cluster  $U$  and  $V$ , respectively, and  $i$  is indexing the feature dimension.

As discussed in section 2.1.2 the different numbers of clusters can exist in parallel because it can be challenging to find the ideal number of clusters for a dataset. Often, the hierarchical clustering is be graphically represented as a tree called *dendrogram* where each node describes a point of split with a parent node and two children nodes (**Figure 2.5right**). The two children nodes are the two clusters that are either merged together in agglomerative clustering or into which the parent cluster is split in divisive clustering. At each level of the hierarchical model, the observations fall into one particular set of clusters where the samples within one group are more similar to each other than between groups. Cutting a dendrogram horizontally at a particular height creates this set of clusters for the dataset. It is split into groups using the vertical lines (**Figure 2.5right**). Interpreting a clustering should be done carefully. Small changes in the dataset can lead to big changes in the dendrogram and naturally hierarchical methods naturally superimpose a structure of clusters onto the dataset irrespective of if such a structure actually exists or not.

## 2.2.4 Visualizations in low dimensional space

The most commonly used statistical technique to reduce the number of dimensions while keeping most of the important information in a dataset is principal component analysis (PCA) [302]. This technique finds a set of uncorrelated vari-

## 2. BACKGROUND

---

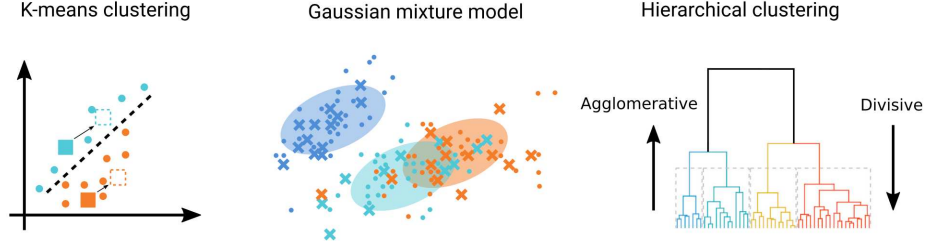


Figure 2.5: **Schematic overview of clustering algorithms.** The three most common statistical techniques to cluster data into groups are K-means, Gaussian mixture model, and hierarchical clustering. For K-means (*left*), points are assigned to the nearest of two centroids and the centroids are iteratively updated until convergence. In Gaussian Mixture Model (*middle*), each point has a probability of belonging to multiple clusters. The method models the data as a mixture of Gaussian distributions, allowing for more flexible cluster shapes. In this example, there are three clusters represented by three Gaussian distributions shown as transparent ellipses. For hierarchical clustering (*right*), clusters are merge or split at different levels of similarity, as visible in the dendrogram structure. For agglomerative clustering, individual data points start in their own clusters and are progressively merged (indicated by arrow from *bottom* to *top*), while for divisive clustering all data points start in one cluster and are recursively divided (indicated by arrow from *top* to *bottom*). The final number of clusters is influenced by the level at which the cut is made (indicated by the *gray* dashed line). In this example, the clustering results in three clusters.

ables called principal components. They are ordered by the amount of variance they explain in the data, with the first component accounting for the most variance. To identify the components, first, the covariance matrix  $\Sigma$  is computed:

$$\Sigma = \frac{1}{n} X^T X \quad (2.21)$$

where  $n$  is the number of data points. Next, the eigenvalue equation is solved to perform eigen decomposition:

$$\Sigma v = \lambda v \quad (2.22)$$

where  $\lambda$  are the eigenvalues and  $v$  are the eigenvectors. The eigenvectors that correspond to the largest eigenvalues are the identified principal components. As the last step, the data is projected onto the principal components:

$$Y = XV \quad (2.23)$$

with  $V$  being the matrix of eigenvectors (each column is an eigenvector). The amount of variance explained by each component is proportional to its corresponding eigenvalue:

$$\text{Variance Explained}_i = \frac{\lambda_i}{\sum \lambda_j} \quad (2.24)$$

Another commonly used statistical technique to reduce dimensionality and to visualize data is t-distributed stochastic neighbour embedding (tSNE), [200]). Here, data points, which often exist in high-dimensional space, are projected into low-dimensional space, with each dot representing one neuron  $x_j$ . The distances between two points  $x_j$  and  $x_i$  reflect the similarities given by conditional probabilities in the high-dimensional space  $p_{j|i}$  and the low-dimensional space  $q_{j|i}$ .

$$p_{j|i} = \frac{\exp(-||x_i - x_j||^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-||x_i - x_k||^2 / 2\sigma_i^2)} \quad (2.25)$$

$$q_{j|i} = \frac{(1 + ||y_i - y_j||^2)^{-1}}{\sum_{k \neq i} (1 + ||y_k - y_l||^2)^{-1}} \quad (2.26)$$

The core of the t-SNE algorithm involves matching the pairwise similarities of the points in the high-dimensional space to those in the low-dimensional space. This is done by minimizing the divergence between two probability distributions using the Kullback-Leibler divergence:

$$C = KL(P||Q) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}} \quad (2.27)$$

This approach is particularly effective in managing the often high-dimensional data encountered in neuroscience.

## 3 | RESULTS

This paragraph will provide the main findings of each project. The results presented are related to:

**Study I** — The combination of stimulus-driven and modulatory inputs in visual thalamus depend on visual responsiveness and stimulus type.

**Study II** — Low-firing, bursting patterns in lateral habenula neurons are linked to behavioral modulation.

**Study III** — Plasticity of neuronal dynamics in the lateral habenula for cue-punishment associative learning.

**Study IV** — Juxtacellular recordings from identified neurons in the mouse locus coeruleus.

### 3.1 Study I: The combination of stimulus-driven and modulatory inputs in visual thalamus depend on visual responsiveness and stimulus type

Preprint available: Schmors\*, L., Kotkat\*, A., Bauer\*, Y., Huang, Z., Crombie, D., Meyerolbersleben, L., Sokoloski, S., Berens, P. and Busse, L. (2023). bioRxiv. <https://doi.org/10.1101/2023.10.18.562960>

\* These authors contributed equally

#### Author contributions

Contributions according to the CRediT taxonomy [8]:

**Conceptualization:** L.B., P.B., S.S.; **Data Curation:** L.S. (data preprocessing, create & maintain research data & software code), A.K. & Y.B. (annotate & produce metadata, preprocess data, maintain research data, maintain software code), L.M., D.C. (pupil size software); **Formal Analysis:** L.S. (statistical analysis, designing and implementing models, decoder), A.K. (statistical analysis, clustering, models for 'blank' conditions), Y.B. (statistical analysis, model refinement); **Funding Acquisition:** L.B., P.B.; **Investigation:** Y.B., A.K., L.M.; **Methodology:** L.S., Y.B., A.K., Z.H., D.C. P.B., S.S., L.B.; **Project Administration:** L.B., P.B.; **Resources:** L.B., P.B.; **Software:** L.S., A.K., Y.B., Z.H., D.C.; **Supervision:** L.B., P.B., S.S.; **Visualization:** L.S. (analysis & evaluation of models, decoder), A.K. (clustering, models for 'blank' condition), Y.B. (schematics, histology, firing rate analysis); **Writing – Original Draft:** L.B., L.S., A.K., Y.B.; **Writing – Review & Editing:** all authors.

#### Introduction

Visual information is processed through a hierarchy of brain areas, which are connected by feedforward and feedback projections. Early in this hierarchy is the dorsolateral geniculate nucleus (dLGN) of the thalamus, a central node for visual information *en route* from the retina to the primary visual cortex (V1) [147, 414]. The dLGN has long been recognized as one of the first visual stages that combines stimulus-driven inputs with additional modulatory inputs [432], such as signals arising from L6 cortico-thalamic (CT) feedback [45, 53, 291, 360, 368], and signals from the brainstem carrying information related to behavioural state [115, 261, 282, 377] and arousal [89, 272, 377, 432]. Since these factors have often been investigated in separate studies, their combined effects on dLGN responses remain poorly understood. Overall, we lack a quantitative understanding of the relative strengths of these modulatory effects on dLGN responses during wakefulness and how they might interact and depend on stimulus-driven input.

On the one hand, it has been firmly established that, even during wakefulness, dLGN responses are modulated according to the animal's internal [39, 282] and overt behavioural state [39, 66, 389]. For instance, in the mouse, locomotion [24, 115] or pupil-indexed arousal [89, 230, 272, 282, 377] are associated with overall enhancements of firing rates in dLGN, which seem to preferentially affect specific neuronal populations depending on their spatio-temporal feature selectivity [24, 230]. Similar to related findings in the somatosensory system [318], the increase in dLGN firing rates seems to be a necessary condition for the sustained depolarisation of primary visual cortex during active behaviours [282].

On the other hand, little consensus has been achieved for dLGN modulations by CT feedback, where a plethora of previous studies have together highlighted its diverse and potentially stimulus-dependent effects. For instance, given that CT

feedback can sharpen spatial dLGN RFs and increase contextual effects [45, 152, 184, 260, 277, 291], CT feedback might enhance or suppress overall firing rates, depending on the size, feature-selectivity and retinotopic position of dLGN neurons [45, 277, 291]. The combination of enhancing and suppressing effects of CT feedback are likely mediated by a differential engagement of both direct excitatory and indirect inhibitory pathways, whose balance will depend on the stimulus selectivity, connectivity and intrinsic properties of corticothalamic neurons in L6 V1, neurons in the thalamic reticular nucleus (TRN) and dLGN [88, 360].

Generalized linear models (GLMs) provide an established framework for statistical analysis of neural responses, that can help to disentangle the combined impact of multiple stimulus-driven and modulatory influences and investigate their properties [62, 297, 309]. While GLMs are relatively simple phenomenological models, they offer the advantage of being interpretable: for instance, GLM kernels learned for the visual stimulus approximate the integration by the spatio-temporal receptive field (RF), and kernels learned for any additional inputs represent spike-induced gain adjustments [298, 309]. First applied in the retina [308, 309], GLMs have since then been used in numerous studies to separate influences of the visual stimulus and other variables, like spike history, interneuronal interaction effects, task-engagement, learning, reward prediction, task-related motor action, locomotion, and arousal [25, 61, 138, 278, 300, 331, 381]. One recent extension of classical GLMs is to estimate RFs by choosing a set of cubic spline basis functions in order to encode smoothness, decrease the number of parameters, and thus be more data efficient [174].

Here, we investigated how feedforward, stimulus-driven signals, feedback signals, and behavioural state jointly influence dLGN activity in awake, head-fixed mice viewing a rich movie stimulus. We simultaneously recorded extracellular dLGN activity, mouse run speed and pupil size, while photosuppressing CT feedback.

### 3. RESULTS

---

We then fitted a spline-GLM model containing kernels for the spatio-temporal RF, CT feedback and behaviour to predict responses of dLGN neurons. The learned kernels were biologically plausible, including diverse spatio-temporal RFs, as well as kernels for behaviour and CT feedback. We found that our model could successfully capture response components derived from the movie input, yielding RF kernels which were generally aligned with response features obtained from simple luminance steps. Including modulatory inputs overall improved the prediction of dLGN responses; the improvements, however, were most prominent for a subpopulation of neurons that were poorly predicted by the movie stimulus. Focusing on effects of CT feedback, we found that these effects depended on stimulus type in both the model and the data, being stronger, more prevalent and faster during the absence of a patterned visual stimulus. Finally, we used the spline-GLM for *in silico* experiments isolating the impact of CT feedback suppression, and demonstrate that stimulus discrimination for CT feedback-modulated neurons was enhanced. We conclude that effects of modulatory inputs in dLGN depend on visual responsiveness and stimulus type.

## Results

### 3.1.1 dLGN responses to movies are modulated by behavioural variables and CT feedback suppression

To investigate how CT feedback, locomotion and arousal modulate thalamic responses, we recorded *in vivo* extracellular dLGN activity in response to a rich movie stimulus in four head-fixed mice together with running speed and pupil size, while randomly photo-suppressing CT feedback ([Figure 3.1a](#)). For photo-suppression of CT feedback, we conditionally expressed the soma-targeting, chloride-conducting channelrhodopsin stGtACR2-RFP [241] in L6 CT pyramidal cells,

by injecting a small volume of Cre-dependent AAV into V1 of Ntsr1-Cre mice [140]. The localisation of stGtACR2 to L6 CT somata and the accurate placement of electrodes were confirmed through post-mortem histological analyses (**Figure 3.1b**). During electrophysiological recordings, the mouse viewed a rich movie stimulus that consisted of a sequence of black-and-white clips from various feature films (‘movies’, **Figure 3.1c, top**). Here, photo-suppression occurred with 50% probability in each 1 second time bin. We also measured the mouse’s run speed and pupil size to infer the animal’s changing behavioural state.

To develop initial insights into the modulations of dLGN responses by CT feedback and the other potentially modulating inputs, we aligned movie responses to onsets of photo-suppression, running, and pupil dilation. We found that certain neurons responded to the onset of CT feedback suppression with a substantial reduction in firing rate (**Figure 3.1d,e**, OFF-ON transitions in *blue*), while others showed milder effects or no modulation at all (**Figure 3.1f, left**). Despite the relatively small effect size, the overall reduction of dLGN firing rates during CT feedback suppression was genuine, as none of the recorded neurons in a control mouse without opsin expression showed systematic modulations at light onset (**Figure A.1a, b**). Furthermore, neural responses aligned to time points in which the light was not switched on (i.e. OFF-OFF transitions in *grey*, **Figure 3.1d, e**) did not show any systematic modulation (**Figure A.1c**). Finally, neurons with stronger CT feedback effects were closer to each other (**Figure A.2**), as predicted by the topography of the corticothalamic system [45, 176, 197, 411]. Consistent with previous findings [24, 115, 229, 377], we also found that during our movie stimulus, firing rates could gradually increase around transitions from sitting to running (**Figure A.3a**), and during pupil dilation (**Figure A.3b**).

These modulations of dLGN responses were also observed across the population of recorded neurons. Specifically, dLGN firing rates during the time of CT feed-

### 3. RESULTS

---

back suppression compared to periods without CT feedback suppression were reduced (*control* vs. *CT FB supp.* mean: 12.6 vs. 11.5 Hz,  $p = 1.24 \times 10^{-2}$ , paired Wilcoxon signed-rank test; **Figure 3.1f, left**), while time windows with running and dilated pupil were associated with an overall increase in average firing rates (*sit* vs. *run* mean: 11.0 vs. 12.8 Hz,  $p = 8.31 \times 10^{-16}$ , **Figure 3.1f, middle**; *small* vs. *large pupil* mean: 11.0 vs. 13.0 Hz,  $p = 4.98 \times 10^{-7}$ , paired Wilcoxon signed-rank test; **Figure 3.1f, right**). Although these modulations affected the recorded dLGN population on average, we also noticed considerable neuron-to-neuron variability. Indeed, for all three modulatory inputs, we found neurons that exhibited substantial variations in firing rates across the different conditions (**Figure 3.1f, insets**).

So far, it seems that our movie stimulus elicited various responses in dLGN, which were modulated by multiple additional inputs, this simple analysis on the mean firing rates does not take into account potential correlations between the different inputs. Consistent with previous studies that have shown that pupil size and running index partially overlapping behavioural states [282, 323, 416], we found a positive correlation between pupil diameter and running speed ( $r = 0.18 \pm 0.17$ , mean  $\pm$  SD;  $p < 0.001$  for 7/10 experiments, permutation test; **Figure A.3c**). Moreover, we found pupil size to also be influenced by stimulus brightness, where lower average intensity of movie frames was associated with larger pupil diameters ( $r = -0.41 \pm 0.15$ , mean  $\pm$  SD;  $p < 0.001$  for 10/10 experiments, permutation test; **Figure A.3c**). Pupil diameter, in turn, is known, to influence responses in the early visual system [44, 129].

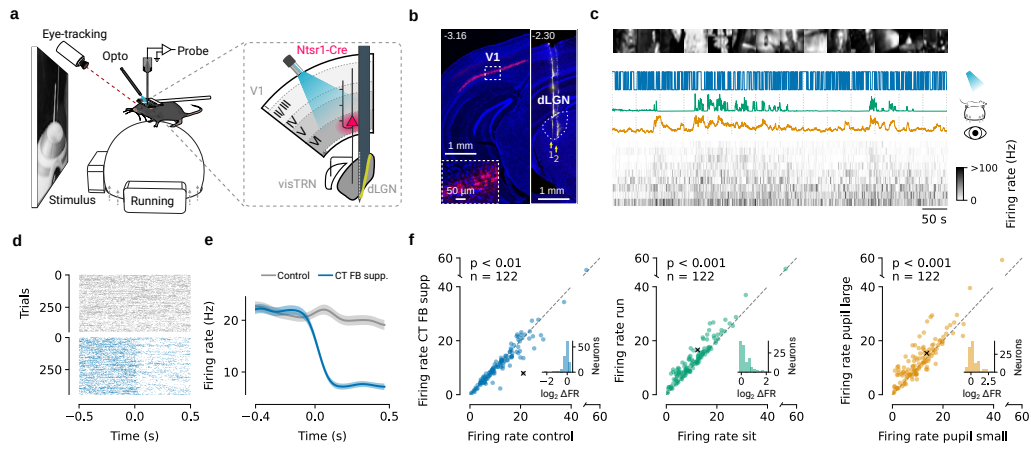


**Figure 3.1: CT feedback and behaviour modulates dLGN responses to movies.** (a) Schematic of the recording setup and photo-suppression of V1 L6 CT pyramidal neurons in Ntsr1-Cre mice with Cre-dependent AAV-stGtACR2-RFP. (b) Histology. *Left*: Coronal section near the V1 injection site, with stGtACR2-RFP expression (red) in Ntsr1+ somata. Blue: DAPI; scale bar: 1 mm. *Inset*: Magnification of area marked by dotted rectangle. Scale bar: 50  $\mu$ m. *Right*: Coronal section of dLGN recording sites, with electrode tracks for two consecutive recording sessions (arrows 1 and 2) marked with DiI (yellow). Scale bar: 1 mm. *Dotted line*: dLGN contour. *Numbers on top*: position relative to Bregma in mm. (c) Snippet of an example dLGN recording. *Top to bottom*: example frames of the movie stimulus, photo-suppression pulse train (blue), running speed (green), pupil area (yellow), and time-varying firing rate (sorted by first principal component) of simultaneously recorded dLGN neurons. (d) Raster plot of responses of an example dLGN neuron, time locked to the onset of CT feedback photo-suppression (blue, OFF-ON transition) and to control periods without photosuppression (grey, OFF-OFF transition). Note that the example neuron illustrates the observed effect of CT feedback photo-suppression, but the size of the effect is not representative of that observed in the population of recorded dLGN neurons. (e) Corresponding PSTHs (solid line: average across trials, shaded area: standard error of the mean). (f) Effects of CT feedback photo-suppression (left), locomotion (middle), and pupil size (right) on dLGN mean firing rates. Example neuron from (d, e) marked with  $\times$ .  $p$  values denote results of a Wilcoxon signed-rank test,  $n = 122$  neurons. *Insets*: Histogram of firing rate fold-change relative to control ( $\Delta$  FR  $\log_2$ -ratio).

#### 3.1.2 A spline-based GLM captures dLGN spatio-temporal RFs and their modulation by CT feedback and behaviour

To disentangle how the various potential influences shape the responses of dLGN neurons, we used a generalized linear model (GLM) [309] that predicted the neuron’s firing rate based on a combination of stimulus-driven and modulatory inputs (running speed, pupil size, and CT feedback suppression) (**Figure 3.2**). The model consisted of one linear kernel for each input, followed by a softplus function that accounted for response nonlinearities (**Figure 3.2a**). The shape of the stimulus kernel captured the neuron’s spatio-temporal RF, while the shapes of the modulatory kernels captured modulations of the neuron’s firing. We employed a GLM with a spline basis [174] in order to efficiently generate smooth kernels, rather than operating directly on the pixels of the visual stimulus or the discrete time bins of the additional inputs (**Figure A.6a–d**). The GLM allowed us to effectively capture the temporal correlations between the inputs (see also above, **Figure A.3c**), and the spatio-temporal correlations in pixel intensities in naturalistic stimuli [121, 301] (**Figure A.4a**).

We trained and evaluated the spline-GLM on the recorded data set as follows: Given the known diversity of mouse dLGN feature selectivity [313, 326], we performed a separate hyperparameter search for each neuron. Hyperparameters included the number of spline bases for the stimulus and modulatory inputs and the weight for L1 regularisation (see Methods). All GLM fits in this study were cross-validated. The models analysed here were configured with the selected optimal hyperparameters and the reported performance is based on the held-out test set.

After model fitting, we assessed the predictive power of the learned kernel shapes using a session-based permutation test (**Figure 3.2b**). To keep the temporal statistics of the time series data intact, we provided the model with input data recorded

on a different day (for the model inputs ‘stimulus’, ‘running’, ‘pupil size’) or with synthetic inputs generated with the same statistics as the original (for the ‘CT feedback suppression’ input). Subsequently, we compared for each input the actual model performance (Pearson’s  $r$ ) against a distribution of model performances with that specific input permuted. Inputs were considered significant if the actual performance differed from the permuted performance with  $p \leq 0.05$ .

We observed a rich diversity of learned GLM kernels for the dLGN neurons’ spatio-temporal RFs, the effects of CT feedback, and the behavioural variables. To begin with, we considered three example neurons (**Figure 3.2c**). Of these, the first had a negative spatial and a transient temporal stimulus response kernel, which contributed significantly to the model’s performance ( $p = 3.51 \times 10^{-2}$ , permutation test; **Figure 3.2c<sub>1</sub>, top**, same neuron as in **Figure 3.1d,e**). In addition, it had a significant negative kernel for CT feedback suppression ( $p = 2.67 \times 10^{-4}$ , permutation test; **Figure 3.2c<sub>2</sub>, c<sub>3</sub>, top**). In contrast, the running and pupil size kernels had minimal impact on this neuron’s model predictions ( $p > 0.05$ , permutation test). The second example neuron exhibited a positive spatial RF with an antagonistic surround ( $p = 2.31 \times 10^{-3}$ , permutation test; **Figure 3.2c<sub>1</sub>, middle**). While both behavioural kernels seemed to contribute (**Figure 3.2c<sub>2</sub>, middle**), only the kernel for pupil size reached significance in the permutation test ( $p = 1.13 \times 10^{-2}$ ). This neuron was also not influenced by CT feedback suppression ( $p > 0.05$ ; **Figure 3.2c<sub>3</sub>, middle**). Finally, example neuron 3 was primarily visually driven, with a positive RF centre and a more sustained temporal response kernel ( $p = 2.31 \times 10^{-3}$ , permutation test; **Figure 3.2c<sub>1</sub>, bottom**). It also had a significant negative kernel for CT feedback suppression ( $p = 2.67 \times 10^{-4}$ , permutation test; **Figure 3.2c<sub>3</sub>, bottom**). The other two modulatory kernels had only negligible influences (permutation test,  $p > 0.05$ ; **Figure 3.2c<sub>3</sub>, bottom**).

The diversity observed in the three example neurons was also evident in the model

### 3. RESULTS

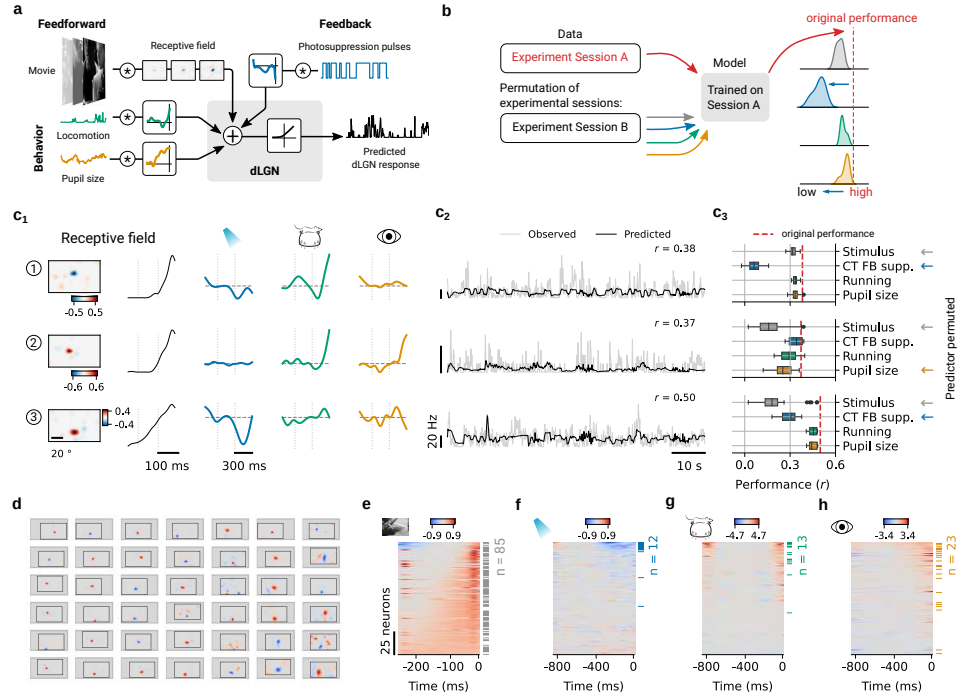


**Figure 3.2: Spline-based GLM captured the RF, the influence of CT feedback and the impact of behaviour on responses of dLGN neurons.** (a) Schematics of the spline-GLM model architecture. Firing rate in dLGN was predicted as a combination of kernel outputs summed at the linear stage and then passed through a softplus nonlinearity. Each modelled neuron had a kernel for the stimulus and kernels for the three modulatory inputs: run speed (green), pupil size (orange), and CT feedback suppression (blue). (b) Schematics of the permutation test [149] to evaluate the significance of the learned kernels. Model performance was evaluated by comparing the actual correlation (Pearson  $r$ ) between predicted and observed firing rates to correlations when one of the inputs was taken from an unrelated experimental session (for movie, running, and pupil size) or randomly generated with the same statistics (for CT feedback suppression). (c) Three dLGN example neurons, their learned kernels, firing rate predictions, and outcomes of the permutation test (neuron 1 is the same example neuron as in ??d-e). (c<sub>1</sub>) Spatial and temporal RF components separated by singular value decomposition (SVD, see Methods), along with kernels for the modulatory inputs. (c<sub>2</sub>) Observed (gray) versus predicted (black) firing rates during 80 s of movie presentation. (c<sub>3</sub>) Actual model performance (Pearson's  $r$ , red dashed line) and performance for permuted stimulus (gray), CT feedback suppression (blue), running (green), and pupil size (orange) inputs. Kernels that contribute significantly to the model's performance are marked with ←. (d) Spatial RFs of example neurons with significant stimulus kernels. Gray: outline of common visual space (azimuth: -35–110 deg; elevation: -35–50 deg); Solid lines: monitor border. (e) Temporal RFs (SVD component of the stimulus kernel) in the recorded dLGN population, sorted by their area under the curve. (f–h) Modulatory kernels in the recorded dLGN population, sorted by their area under the curve, for CT feedback suppression (f), running (g), and pupil size (h). Horizontal bars, side: Neurons with significant kernels based on the permutation test. Panels (e–h) show data from all  $n = 122$  neurons.

fits across the population of recorded dLGN neurons, where we obtained a variety of spatio-temporal RFs, and combinations of modulatory influences. Assessing the learned spatio-temporal kernels, we found that  $\sim 70\%$  of the recorded dLGN neurons were visually responsive to our movie stimulus (85/122 neurons; permutation test visual stimulus,  $p \leq 0.05$ ; [Figure A.6e](#)). For many of these neurons (representative examples in [Figure 3.2d](#)), the spline-GLM recovered spatial RF properties that were consistent with previous descriptions of mouse dLGN RFs obtained using artificial stimuli [[112](#), [313](#), [396](#)] and in many cases resembled those obtained from conventional sparse noise experiments ([Figure A.4b,c](#)). These properties included various RF locations, RFs with either positive (66%, 56/85 neurons) or negative polarity kernels (34%, 29/85 neurons), a broad range of RF surround strengths ([Figure A.4d,f](#)), and various RF centre sizes ([Figure A.4e](#)). Furthermore, the GLM captured the well-known diversity of temporal response properties of dLGN neurons [[145](#), [313](#)], with some neurons showing more sustained, and others showing more transient temporal kernels ([Figure 3.2e](#)). Importantly, both the spatial and temporal GLM kernels matched well with the expected polarity and the dynamics obtained from clustering responses into sustained-OFF, Sustained-ON, and Transient groups during full-field luminance steps ([Figure A.5](#)). This correspondence further underscores our model’s capacity to capture meaningful spatio-temporal RF properties and essential visual response characteristics of the recorded dLGN neurons.

We next assessed the learned kernels for the modulatory inputs. According to the permutation test, approximately 10% (12/122 neurons) of the recorded dLGN neurons were affected by CT feedback suppression, 11% (13/122 neurons) by running, and 19% (23/122 neurons) by pupil size ([Figure 3.2f–h](#), [Figure A.6f–h](#)). The overall direction of modulation for the significant kernels aligned with the modulation indices obtained directly from the data ([Figure 3.1f](#)): for the sig-

### 3. RESULTS

---

nificantly modulated neurons, CT feedback suppression kernels were predominantly negative (**Figure 3.2f**), while running speed (**Figure 3.2g**) and pupil size (**Figure 3.2h**) kernels were predominantly positive. These results indicate that both CT feedback and behavioural state variables during natural movie viewing contribute to an overall increase of dLGN responses. Beyond the general sign of modulation, the learned model kernels also offered insights into the temporal dynamics of the modulatory influences: consistent with the well-known slow impact of pupil indexed arousal on responses in the visual system [323, 416], we found that kernels for pupil size had a more sustained profile compared to kernels for running modulations (time to half max kernel running vs. pupil size: 106.3 ms vs. 146.3 ms, Wilcoxon signed-rank test:  $p = 7.51 \times 10^{-3}$ , **Figure A.6p**). This ability to learn differential kernels and a close correspondence between data-driven modulation indices and the impact of the modulatory inputs for model performance (**Figure A.7d–g**) demonstrates that our GLM model was successful in extracting the impact of the various modulatory inputs.

Despite the spline-GLM’s overall success in predicting dLGN responses and learning biologically plausible kernels, it faced challenges in capturing fast modulations and response peaks, as typical also for traditional GLMs [293, 297]. We also observed increased variance at the left side of some kernels particularly evident for running modulated neurons (**Figure 3.2e–h** and **Figure A.6i–l**). This could potentially stem from typical boundary artefacts associated with splines (**Figure A.6a–d**). Nevertheless, despite these challenges, incorporating running information proved beneficial, as confirmed by the permutation test (**Figure A.6g**). Reassuringly, these running modulated neurons showed reduced response reliability in repeated stimulus trials (**Figure A.6m,n**) as their activity was substantially influenced by locomotion state. In line with a previous study [97], these neurons also had overall low firing rates (**Figure A.6o**).

Is the contribution of modulatory inputs to dLGN neurons' activity consistent and strong enough to improve the prediction of dLGN responses? While dLGN has long been known to exhibit state-dependent changes in firing [24, 39, 115, 287] and to receive extensive feedback from cortex [52, 367], the impact of these influences, in particular during viewing of naturalistic stimuli, is not well understood. Thus, to quantitatively assess the contributions of CT feedback suppression and behavioural variables, we compared the performance between our model including all inputs and reduced variants of the model with only a subset of inputs (**Figure 3.3a,b**). Starting with the 'Stimulus only' model that only considered the stimulus as input, we found that incorporating one or more modulatory inputs increased the correlation between observed and predicted dLGN responses (ANOVA:  $p = 0.022$ ; **Figure 3.3a**). In particular, adding pupil size or a combination of two or more predictors showed a significantly better performance than the 'Stimulus only' model ('Stimulus only' vs. 'Stimulus + Pupil size': 0.186 vs. 0.235, Wilcoxon signed-rank test: Bonferroni corrected  $p = 3.39 \times 10^{-6}$ ; 'Stimulus only' vs. 'Full model': 0.186 vs. 0.249, Wilcoxon signed-rank test: Bonferroni corrected  $p = 2.08 \times 10^{-7}$ ; **Figure 3.3a**). Note that even in the full model we still observed neurons with suboptimal predictions, maybe due to low response reliability (**Figure A.6m,n**) and sparse firing (**Figure A.6o**).

Indeed, when we compared the prediction performance of the 'Stimulus-only' model with the 'Full model' (**Figure 3.3b**), we noticed that the inclusion of the modulatory inputs did not merely shift the distribution to higher performances. Instead, a closer examination allowed us to identify a subgroup of neurons whose responses could be predicted only poorly by the visual stimulus, but which showed substantial improvements in their response prediction with the inclusion of modulatory factors (**Figure 3.3c,d**; we call them the 'Modulation-sensitive' group; **Figure A.7a-f**). In contrast, another subgroup showed no improvement when

### 3. RESULTS

---

we added the modulatory inputs (**Figure 3.3c,d**; 'Stimulus-explained' group). To test the extent to which this reflected a ceiling effect in explainable variance, we fitted models that had only the modulatory factor as inputs but not the stimulus ('CT FB supp.' + 'Running' + 'Pupil size'; 'Modulation only' models; **Figure 3.3f**). For these models, neurons substantially affected by modulatory inputs still performed well (**Figure 3.3f**, 'Modulation-sensitive', dark dots), whereas those well-explained by the visual stimulus exhibited lower performance, although not zero (**Figure 3.3f**, 'Stimulus-explained', light grey dots). This indicates that even in the 'Stimulus-explained' group, modulatory inputs might have some, albeit generally weak effect that was obscured in the full model; alternatively, some modulatory factors might contain stimulus information, in particular pupil size, which is known to be related not only to arousal, but also to stimulus brightness (**Figure A.3c**). Taken together, our analysis of model performance suggests that dLGN neurons are explained by the visual stimulus, behaviour and CT feedback, albeit with considerable heterogeneity.

#### 3.1.3 CT feedback is enhanced in the absence of a patterned visual stimulus

Our model showed only a relatively small subset of dLGN neurons with significant effects of CT feedback. Could the reason for this be that the responses elicited by the rich naturalistic movie stimulus might have dominated dLGN activity relative to the effects of CT feedback? We thus predicted that CT feedback effects might be stronger without patterned stimulus input (**Figure 3.4a**).

To test this prediction, we expanded our analyses to the period of blank screen stimulation flanking the movie presentation, and compared the effects of CT feedback suppression in the recorded data and in models fit separately to responses

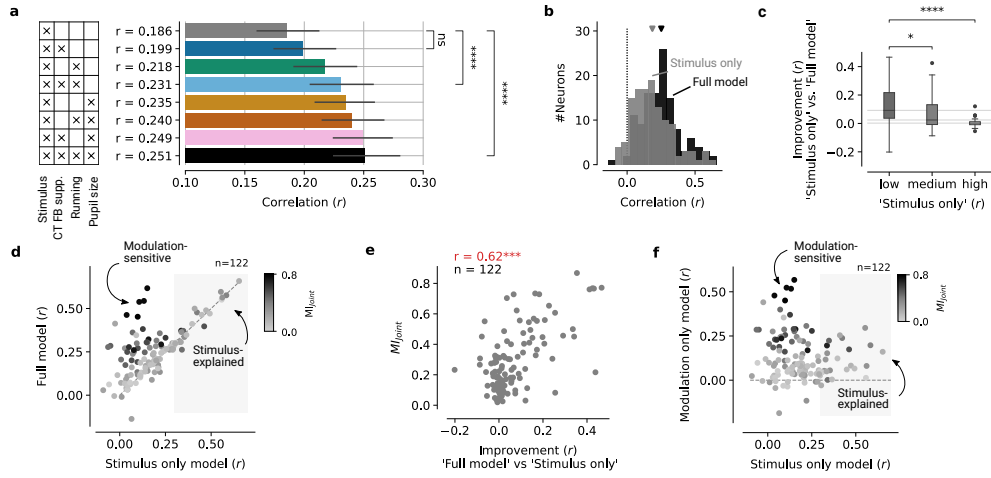


**Figure 3.3: Adding model predictors for modulatory inputs improves the performance for a subgroup of poorly visually responsive dLGN neurons.** **(a)** Comparison of model performance for models with different sets of inputs indicated in the table on the left, sorted by performance. Bonferroni corrected  $p$ -values of paired Wilcoxon signed-rank test:  $**** p \leq 1.0 \times 10^{-4}$ ; 'ns' non-significant. Error bars: 95% confidence intervals. **(b)** Comparison of model performance on the test set (Pearson's  $r$ ) for the 'Full model' (black, inputs: stimulus, CT feedback suppression, run speed, pupil size) and the 'Stimulus only' model (gray). Arrow heads: mean performance ( $n = 122$  neurons). **(c)** Improvement in model performance observed with adding predictors for modulatory inputs ('Full model' – 'Stimulus only') for neurons grouped by the (low/medium/high) performance of the 'Stimulus only' model. Bonferroni corrected  $p$ -values of paired Wilcoxon test:  $*p \leq 0.05$ ;  $****p \leq 1.0 \times 10^{-4}$ . **(d)** Comparison of model performances on the test set for all neurons in the 'Stimulus only' model and the 'Full model'. Darker colours indicate stronger joint modulation by CT feedback, running, and pupil size ( $MI_{\text{Joint}}$ ) estimated directly from the data without model (see Methods). Arrows indicate the group of neurons that benefits from adding CT feedback, running, and pupil size ('Modulation-explained') and the group that does not improve ('Stimulus-explained'). **(e)** Improvement in model performance observed with adding predictors for modulatory inputs ('Full model' – 'Stimulus only') and the joint modulation by CT feedback, running, and pupil size estimated directly from the data without model (see Methods). **(f)** Performances of a model that considers only the stimulus as input ('Stimulus only') and a model without the stimulus ('Modulation only'). Arrows indicate 'Modulation-explained' and 'Stimulus-explained' neurons. Darker colours indicate stronger joint modulation by CT feedback, running, and pupil size estimated directly from the data without model. Panels (d–f) show data from all  $n = 122$  neurons.

### 3. RESULTS

---

during movies vs. blank periods (**Figure 3.4b**). Consistent with our prediction, we indeed observed that CT feedback suppression reduced the firing rates of individual example neurons more strongly during blanks ( $MI_{CT\ FB\ supp.}$ : -0.65, **Figure 3.4c<sub>1</sub>**) compared to movies ( $MI_{CT\ FB\ supp.}$ : -0.45, **Figure 3.4c<sub>2</sub>**). This observation held true across the recorded dLGN population (mean  $MI_{CT\ FB\ supp.}$ : -0.09 vs. -0.03;  $p = 2.5 \times 10^{-6}$ , Wilcoxon signed-rank test; **Figure 3.4d**). Furthermore, during blank periods, more dLGN neurons were affected by the suppression of CT feedback (45% vs. 18% with  $|MI_{CT\ FB\ supp.}| \geq 0.1$ ; **Figure 3.4j**). The stronger effect of CT feedback suppression could not be explained by the difference in overall firing rate between the two stimulus conditions (**Figure A.8a-b**), nor was it related to the different number of optogenetic pulses (**Figure A.8c**).

Is the stronger modulation by CT feedback suppression during blank periods vs. movies also captured by our spline-GLM model? We indeed found that the model learned a more negative CT feedback suppression kernel for the blank condition compared to the movie, both in the example neuron (peak amplitude -1.8 vs. -0.94; **Figure 3.4f**) as well as in the population of neurons with a significant CT feedback suppression kernel (-0.86 vs. -0.46,  $p = 0.044$ , Mann-Whitney-U test, **Figure 3.4g**; all neurons: -0.4 vs. -0.2,  $p = 5 \times 10^{-8}$ , Wilcoxon signed-rank test). To avoid potential confounds of firing rates on kernel amplitudes (**Figure A.8d,e**) and enable a more direct comparison to the recorded data, we used the model predictions in the two stimulus conditions and calculated model-derived  $MI_{CT\ FB\ supp.}$ , analogously to those based on the recorded data. Consistent with our data-driven observations, we found that CT feedback suppression reduced modelled responses more strongly during the blank periods (modelled  $MI_{CT\ FB\ supp.}$ , -0.057) compared to movies (-0.017;  $p = 0.002$ , Wilcoxon signed-rank test, **Figure 3.4h**).

Inspecting the learned model kernels (**Figure 3.4f,g**) suggested that, besides amplitude, the dynamics of the CT feedback suppression effects might also differ

between visual stimulus conditions. Specifically, the spline-GLM learned a kernel characterised by a faster and briefer time course for blank periods compared to movie stimulation. This observation could be quantified by calculating the rate of change (RoC) of the neurons' predicted and actual responses after photosuppression (see Methods). Indeed, we found that the effect of CT feedback suppression was faster during blank periods (model-based RoC:  $p = 0.004$ , Wilcoxon signed-rank test, [Figure 3.4i](#); data-based:  $p = 9 \times 10^{-12}$ , Wilcoxon signed-rank test, [Figure 3.4j](#)). In conclusion, the fitted spline-GLM models revealed that the observed impact of CT feedback suppression was weaker and slower during the presence of a rich patterned visual stimulus in comparison to blank periods, suggesting that the observed effect of CT feedback depends on the visual stimulus characteristics.

### 3.1.4 CT feedback affects stimulus decoding

Finally, we sought to better understand if and how the suppression of CT feedback affected how dLGN encoded visual information ([Figure 3.5](#)). Specifically, we tested the degree to which suppression of CT feedback might affect the decoding of movie information based on the responses of single dLGN neurons, compared to the control condition where feedback was intact.

To address our question, we used the trained model ('Full model') to describe how individual dLGN neurons encode stimulus information ([Figure 3.5a](#)). In this way, we were able to isolate the effects of CT feedback suppression *in silico*, by keeping stimulus, running and pupil inputs intact, and only changing CT feedback suppression (setting the CT feedback suppression either OFF (0) or ON (1)). We then performed a decoding experiment, using simulated model responses to discriminate between video clips in a two-alternative forced choice (2AFC) setting [308]. We randomly selected 50 ms fragments of stimulus-response-pairs

### 3. RESULTS

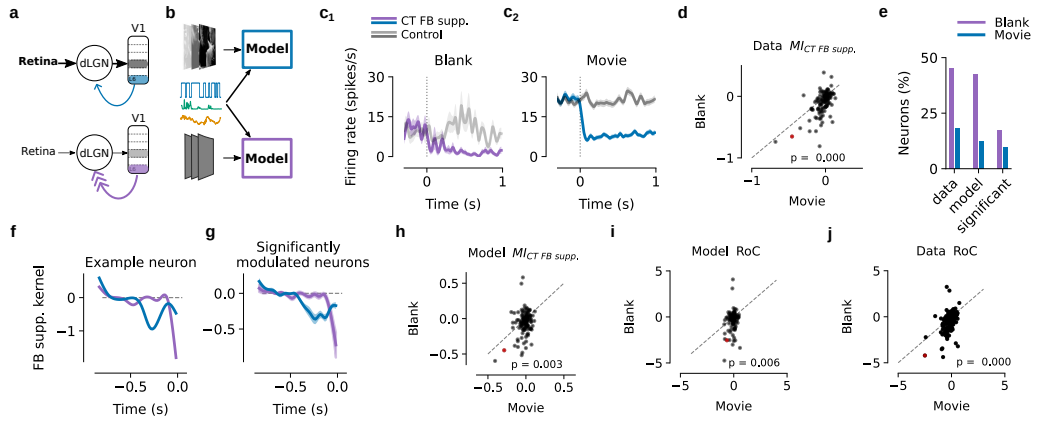


**Figure 3.4: The effect of CT feedback is dependent on the presence or absence of the visual stimulus.** (a) Schematics of the hypothesis that the absence of a patterned visual stimulus elicits weaker stimulus-driven input to dLGN, which is in turn accompanied by stronger CT feedback. Note that the schematics should not imply stronger L6 CT pyramidal neuron firing, but that the net effect of CT feedback on dLGN firing is stronger in the absence of a patterned visual stimulus. (b) Schematics of fitting the spline-GLM model separately during movies (*top*) vs. blank periods (*bottom*). (c) Effects of CT feedback suppression. (c<sub>1</sub>) Mean PSTHs time locked to onset of photosuppression for one example dLGN neuron (same example neuron as in ??d-e) during blank periods (gray screen) flanking the movie presentation. The shaded area represents the SEM. (c<sub>2</sub>) Same as (c<sub>1</sub>), during movie presentation. *Purple, blue*: PSTH during CT feedback suppression (OFF-ON transition), *light grey, dark grey*: PSTH during control condition (OFF-OFF transition). (d) Comparison of  $MI_{CT\ FB\ supp.}$  during blanks vs. movies for the recorded dLGN population (number of neurons  $n = 122$ ). (e) Percentage of recorded dLGN neurons modulated by CT feedback during the two stimulus conditions. Three modulation metrics were separately considered to count the modulated neurons. A neuron was considered modulated (1) based on data:  $|MI_{CT\ FB\ supp.}|$  from (d)  $\geq 0.1$ , (2) based on model predictions:  $|MI_{CT\ FB\ supp.}|$  from (h)  $\geq 0.1$ , or (3) based on model performance: permutation test  $p \leq 0.05$ . Notably, all three modulation metrics consistently revealed a higher proportion of neurons displaying CT feedback modulation during the blank condition compared to the movie condition. (f) CT feedback suppression kernel for the example dLGN neuron in (c<sub>1</sub>). The model was either trained on the data from movie presentation (*dark blue*) or blank periods (*light blue*). (g) Same for all significantly CT feedback modulated neurons ( $n_{movie} = 12$ ,  $n_{blank} = 21$ ). Solid lines represent the mean of the kernels and transparent surrounds represent the standard error of the mean (SEM). (h) Comparison of  $MI_{CT\ FB\ supp.}$  for blanks vs. movies calculated from simulated data using the fitted model for the dLGN population to the two stimuli. (i) Comparison of the Rate of Change (RoC) of model-predicted neurons' responses to CT feedback suppression during movie presentation vs. blank periods. (j) Same as (i), for the recorded data. Panels (h–j) show data from all 122 neurons.

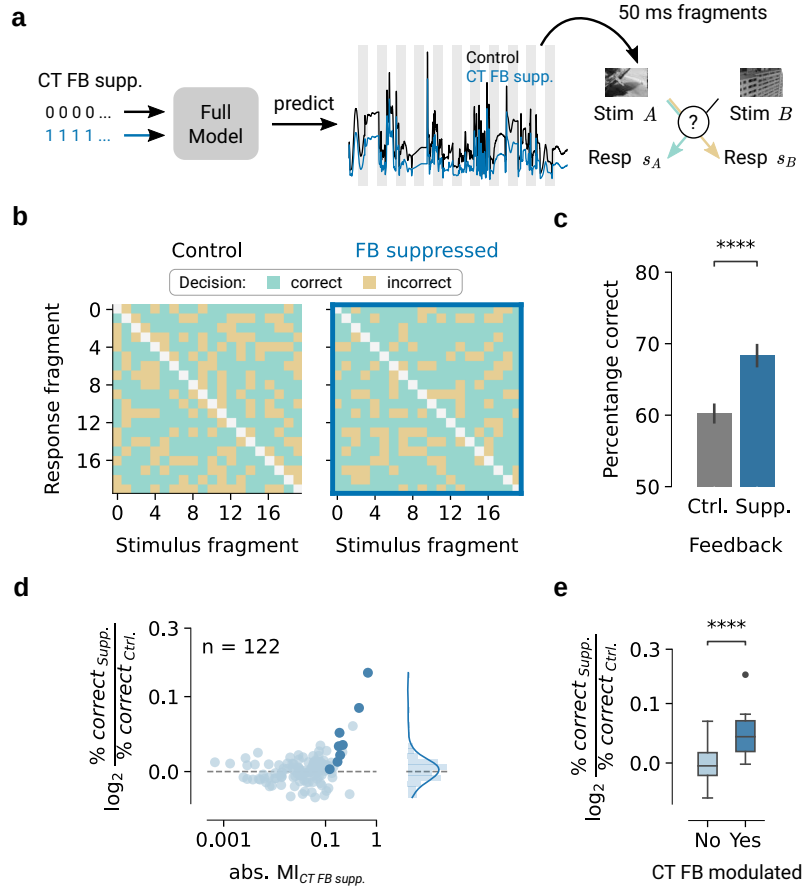


**Figure 3.5: Two-alternative-forced-choice decoder shows better stimulus discrimination for CT feedback modulated neurons when feedback is suppressed.** (a) Schematics of the decoder: the trained model was used to simulate responses with CT feedback being either on (blue) or off (black; 70 s example trace). Random 50 ms stimulus fragments (*A* and *B*) were chosen, each with corresponding simulated responses. The decoder's task was to determine which stimulus was more likely based on the observed responses. (b) Decision matrix for one example neuron with 20 random stimulus fragments and their simulated responses. Green: correct pairing of stimulus and response; yellow: incorrect pairing. Left: control condition; right: CT feedback suppressed condition. (c) Percentage of correct decisions in the control and the CT feedback suppressed condition, same example modelled neuron as in (b). Error bars indicate 95% confidence intervals. (d) Decoder performance across all neurons ( $n = 122$ ) as a function of the amount of modulation by photosuppression observed in the experiment. Dark blue: neurons with a significant CT feedback kernel. (e) Relative decoder performance during CT feedback suppression and control conditions, split according to whether neurons were significantly modulated by CT feedback suppression (dark blue) or not modulated (light blue).

### 3. RESULTS

---

from the test set that had not been used for model fitting. To discriminate between two movie clips, we used the maximum-likelihood decoding rule from Pillow et al. (2005) [308]. The decision of the 2AFC decoder was directly derived from the model likelihood for the correct and incorrect pairing (see Methods and [308]). We considered the 2AFC decoder to make a correct decision if the likelihood for the correct pairing was higher compared to the incorrect one. We found that the decoder made more correct decisions in the FB suppressed condition compared to the control condition (**Figure 3.5b**, 20 representative stimulus-response pairs for one example neuron; **Figure 3.5c**, across all 100 stimulus pairs 68.34 % vs. 60.27 %,  $p = 1.20 \times 10^{-12}$ , Wilcoxon rank sum test).

To quantify the discrimination performance for each neuron, we computed the log ratio of the percent of correct choices in the FB suppressed condition and the control condition (**Figure 3.5d**), with positive log ratios indicating better decoding performance in the CT feedback suppressed state. The observed log-ratios near zero for the majority of neurons suggest that there was no difference in stimulus information between control conditions and CT feedback suppression (average  $1.003 \pm 0.021$ , mean  $\pm$  SD; **Figure 3.5d**). However, we found significantly higher log-ratios in neurons modulated by CT feedback compared to non-modulated neurons ( $p = 5.84 \times 10^{-14}$ , skew test against null; **Figure 3.5d**, **right** and **Figure 3.5e**). Consequently, our model predicts that with intact CT feedback, short duration activity of individual dLGN neurons can be less informative about the visual stimulus.

### Discussion

Here, we quantitatively characterised how dLGN responses during viewing of a naturalistic movie are influenced by the combination of visual stimulus-related inputs, CT feedback and behavioural state. We modelled the responses of individual

dLGN neurons using a data-efficient GLM, which predicted the spatio-temporal RFs and response modulations by CT FB suppression, the animal’s run speed, and pupil size. We found that overall model performance improved when including the modulatory predictors, in particular for a subpopulation of dLGN neurons whose responses were poorly explained by the visual stimulus. Guided by the model, we found that the effect of CT FB suppression depended on stimulus type, being relatively stronger, more prevalent, and faster in the absence of a patterned visual stimulus. Finally, using Bayesian decoding, we found that the responses of a subset of CT-feedback-sensitive dLGN neurons contained more visual information when CT feedback was suppressed. Together, our results show that the activity of dLGN neurons is governed by a combination of stimulus-driven and modulatory inputs and is mediated by their visual responsiveness and the stimulus type.

### **Using a spline-GLM for modelling thalamic responses**

Our work extends previous approaches that modelled thalamic processing of naturalistic visual stimuli (e.g. [25, 61, 221]). In particular, our framework allowed us to test model variants with different combination of predictors, which revealed a subpopulation of dLGN neurons whose responses were poorly predicted by the movie stimulus and more strongly affected by the modulatory inputs. These might correspond to a previously reported set of neurons in mouse dLGN, amounting to 30–40%, with poor or unclear visual feature selectivity [145, 366]. Given that this subpopulation also contained an over-representation of low-firing neurons, their relatively stronger modulation (see also [97]) might be indicative of a tight inhibitory / excitatory coupling in the recorded network [337]. Such tight coupling in the thalamo-cortico-thalamic network might be mechanistically achieved through joint modulations of dLGN relay cells, and inhibitory neurons

### 3. RESULTS

---

in dLGN and the thalamic reticular nucleus by CT feedback and neuromodulation [45, 147, 282, 368].

The general success of our model is indicated by the close neuron-by-neuron correspondence between RF properties and response types derived from the model and more traditional analyses and stimuli, and by the prominent relationship between learned model kernels and data-driven modulation indices for CT feedback suppression, run speed and pupil size. Yet, while our data-driven modulation indices generally matched well with the magnitudes of modulatory influences previously observed in mouse dLGN [24, 45, 115, 272, 377], we were surprised to find only a relatively small fraction of dLGN neurons with significant kernels for CT feedback suppression, running and pupil size. Reasons for the small fraction might be at least threefold: (1) the conservative method for assessing significance in the model derived from session-based permutation tests [149], (2) the continuous nature of inputs to the model rather than a split into extreme conditions for some of the modulation indices (e.g. locomotion slower than 0.25 cm/s was considered as sitting but only faster than 1 cm/s was considered as running [24, 115]), (3) the direct suppression of L6 CT feedback through the light-gated chloride channel stGtACR2 [241], which might have yielded comparatively weaker effects (see also [324, 377]) than alternative approaches recruiting powerful intracortical inhibition through photostimulation of inhibitory V1 neurons [45, 193, 282, 291, 324, 377].

In the future, our spline-GLM could be extended by thalamic mechanisms, such as the fast adaptation of integration time according to luminance and contrast [221, 244], accounting for the constant changes in spatial and temporal integration elicited by dynamic natural stimuli [81]. Further, by combining the model inputs, one could test for interactions, for instance to clarify the dependence of CT feedback effects and behavioural state-related modulations as proposed by some stud-

ies [324]. Incorporating adaptive amplitudes of input and post-spike kernels [428] could differentiate tonic and burst spiking behaviours, characteristic of thalamic neurons [361] and known to be affected by both stimulus-related [220] and modulatory inputs [39, 115, 286, 377]. Finally, alternative modelling approaches based on deep neural networks (DNNs) [232, 275, 446] promise to better capture non-linear dynamics and provide faster computations, potentially enhancing our capabilities to emulate neuronal responses and complex neural interactions, yet likely at the expense of interpretability and uncontrolled biases [110, 133]. Some recent efforts have been made to increase the interpretability and biological plausibility of such models [87, 239], at times, by crafting an integrative model combining the advantageous non-linearity of DNNs with linear models like ours [199, 236].

### **Spatio-temporal RFs and modulatory inputs of mouse dLGN neurons**

Through a combination of analysing receptive fields obtained from the model using a naturalistic movie stimulus, a noise stimulus, and more traditional RF mapping techniques, we identified both expected and unexpected types of dLGN spatio-temporal RFs. Indeed, consistent with previous studies quantifying RFs in mouse dLGN to simple stimuli [112, 145, 313, 396], our model learned circular spatial RFs for the majority of neurons, which often consisted of a single domain resembling the well known ON and OFF fields. Reminiscent of a study showing that retinal ganglion cells can reverse the polarity of their RFs in response to different natural images [137], we found that some dLGN neurons had an opposite RF polarity when characterised with simple luminance steps or the movie stimulus. Finally, some of the RFs obtained with our modelling approach had a complex spatial structure, and might thus correspond to a subset of dLGN neurons that had been previously noted to lack clearly localised RFs [112, 145]. Future studies, for

### 3. RESULTS

---

instanced based on the “maximally exciting image” approach initially applied to mouse V1 [420], are needed to verify to which degree these complex RF structures indeed reflect complex feature selectivity of dLGN neurons or might be potential consequences of the modelling approach.

Despite its frequent portrayal as a relay station, the dLGN of the thalamus has long been recognised as one of the earliest stages in the visual system that integrates visual representations with additional information [39, 261, 362, 368, 432]. Corroborating this, we found that the performance of our model generally improved with the inclusion of modulatory inputs, albeit with considerable neuron-by-neuron diversity. Previous studies have already reported differential effects of behavioural state and arousal, and could relate them to the neurons’ feature selectivity [24, 230, 268]. In dLGN, the firing of neurons with nonlinear responses to high spatial frequencies and with transient ON responses seem preferentially enhanced by locomotion [24]. In addition, retinal boutons preferring low spatial frequencies and luminance decrements seem preferentially suppressed by arousal [230]. The neuron-by-neuron diversity in the impact of the modulatory inputs might thus serve to enhance particular visual inputs during active states.

#### **CT feedback effects on single dLGN neuron stimulus encoding and decoding**

Our observation of faster and relatively stronger CT feedback effects during blank periods compared to movie viewing contributes to the growing appreciation that CT feedback effects on dLGN firing rates seem to be stimulus-dependent, and potentially overridden by strong visual stimulation (see also [377, 460]). Indeed, the effect of CT feedback seems to be most potent in the absence of patterned stimulus input, both in mice (this study and [197, 377] for related findings with

gratings) and ferrets [460], which might point to a common mechanism across species. We propose that a greater influence of CT feedback suppression during spontaneous activity than movie viewing could arise from a differential engagement of direct excitatory and inhibitory feedback pathways during these visual stimulus types. Supplying thalamic relay neurons with different ratios of excitatory and inhibitory conductances, mimicking the impact of modulatory inputs, can shift their input/output function, such that the same somatic input can generate markedly different spiking responses [33, 440]. Future studies will need to use more subtle manipulations of stimulus type, including contrast and spatial structure, together with pathway-specific CT feedback suppression, to test the hypothesis that CT feedback might be most effective under challenging sensory conditions. In addition, future studies will profit from simultaneous recordings of V1 and dLGN to further disentangle the origin and potential interaction of CT feedback and other modulatory influences.

We also observed that suppressing CT feedback enhanced the decoding of visual information in individual, feedback-sensitive neurons. This was surprising given prior research associating cortico-cortical feedback with attention and enhancements of stimulus encoding (e.g. [157, 209, 325, 400, 456]). What could be potential reasons why our model predicted improved decoding during CT feedback suppression? Past research has proposed that one role of CT feedback could be to linearise the input-output relationship of thalamic neurons through the injection of synaptic noise [33, 440]. This linearisation would enhance the excitability of thalamic neurons to weaker inputs, but would additionally push them into more unreliable firing regimes. This could be one potential explanation why our decoder operating on single stimulus segments performed worse with CT feedback intact. A shift towards more unreliable firing regimes by CT feedback would also be in line with results of a previous study [377], showing that CT feedback during

### 3. RESULTS

---

repeated movie segments increased the trial-by-trial variability of dLGN neurons' responses. Such increased variability in individual dLGN neurons could be offset and even exploited by the strong convergence in the thalamocortical system [10, 415]. Specifically, synaptic noise mediated by CT feedback might allow to extract from the pooled afferent signal stimulus-related information with better resolution and with enhanced sensitivity to weaker inputs [33, 440]. Thus, in the future, a more realistic decoder would consider local populations of thalamocortical neurons, to test the hypothesis that, through CT feedback mediated synaptic noise, cortex regulates the trade-off between heightened sensitivity towards ambiguous inputs and reliable representation of clear inputs.

In conclusion, our results add to the growing body of evidence that dLGN activity is influenced not only by visual inputs but also by modulatory influences from CT feedback and behavioural state. Our work presents an important step towards a quantitative understanding of how dLGN responses to complex, naturalistic stimuli are shaped by the simultaneous influences of stimulus-related feedforward inputs, CT feedback and behaviour.

## 3.2 Study II: Low-firing, bursting patterns in lateral habenula neurons are linked to behavioral modulation.

Manuscript in preparation: Ioannis S. Zouridis <sup>\*</sup>, Lisa Schmors <sup>\*</sup>, Mauro Congiu, Salvatore Lecca, Manuel Mameli, Philipp Berens, Andrea Buraloss

<sup>\*</sup> These authors contributed equally

### Author contributions

Contributions according to the CRediT taxonomy [8]:

**Conceptualization:** I.S.Z., L.S., P.B., A.B.; **Data Curation:** I.S.Z. (annotate & produce metadata, data preprocessing, maintain research data, maintain software code), L.S. (data preprocessing, create & maintain software code), M.M. (produced metadata for data published in Congiou et al. 2023); **Formal Analysis:** I.S.Z. (data processing, statistical analysis), L.S. (data processing, statistical analysis, computational models); **Funding Acquisition:** P.B., A.B.; **Investigation:** I.S.Z. (electrophysiological and behavioural recordings), M.C. & S.L. (contributed electrophysiological data published in Congiou et al. 2023); **Methodology:** I.S.Z. (electrophysiology, behaviour, anatomy experiments), L.S. (creation of computational models); **Project Administration:** I.S.Z., A.B., L.S.; **Resources:** P.B., A.B.; **Software:** L.S. (software development for statistical analysis and computational models), I.S.Z. (software for data preprocessing); **Supervision:** P.B., A.B., M.M.; **Validation:** I.S.Z., L.S.; **Visualization:** L.S. (designed and produced figures), I.S.Z. (contributed designing figures); **Writing – Original Draft:** I.S.Z., L.S.; **Writing – Review & Editing:** all authors.

#### Introduction

The lateral habenula (LHb) is a bilateral epithalamic structure that integrates signals associated with aversive experiences to regulate motivated behaviors [172, 273]. The LHb was initially thought to be a homogenous brain structure, composed almost exclusively of excitatory glutamatergic neurons [6]. However, recent work has revealed a high degree of heterogeneity among LHb neurons in respect to morphological, electrophysiological, molecular and connectivity properties [84]. Understanding this heterogeneity is essential for deciphering how LHb integrates information to influence brain states and behavior [266]. Ex-vivo studies have classified LHb neurons into several morphological and electrophysiological subtypes [419]. The progress in omics technologies has confirmed this heterogeneity at the molecular level [150, 422]. Moreover, transcriptionally defined LHb subtypes have been shown to have distinct connectivity, indicating a modular network architecture [422]. The morphological, molecular and electrophysiological heterogeneity observed ex vivo is also paralleled by a high degree of diversity in-vivo. LHb neurons display diverse in-vivo spiking patterns as well as functional properties: for example, different LHb units fire at different rates and with different temporal spiking patterns [430, 437], they respond differently to aversive stimuli, and engage differently in learning [83, 84]. This diversity of in-vivo activity patterns, and their contribution to the encoding of aversive stimuli and behavior, however, has remained largely unresolved. In this study, we aimed at filling this gap. Specifically, we used a Mixture of Gaussian model on in-vivo electrophysiological data to identify and quantify the diversity of LHb neuronal firing patterns in both anesthetized and awake mice. Moreover, we related spontaneous firing patterns to responses to aversive stimuli and to spontaneous behavioral state fluctuations. By resolving the heterogeneity of LHb neurons and

examine how this relates to their functional responses, we aim to provide a more accurate and detailed classification of LHb neurons and their modulation by aversive stimuli and behavioral states.

## Results

### 3.2.1 LHb neurons spontaneously fire in four distinct patterns

To explore the diversity of spontaneous firing patterns of LHb neurons, we compiled datasets of in vivo juxtacellular recordings from anesthetized and awake preparations (**Figure B.2c,d**), and visualized the firing patterns in two-dimensional space using t-distributed Stochastic Neighbor Embedding [200]. To investigate the number and nature of spontaneous firing patterns of LHb neurons, we used Gaussian mixture model (GMM) clustering on each dataset separately. This analysis yielded four optimal clusters in each dataset, referred to as firing patterns 1 through 4 thereafter (**Figure 3.6a,b**). To compare the firing pattern types of LHb neurons across awake and anesthetized states, we measured the cluster similarity by computing the cosine similarity and matching them using the Hungarian algorithm (see methods; **Figure 3.6c**,  $S_C(1) = 0.76$ ,  $S_C(2) = 0.60$ ,  $S_C(3) = 0.82$ ,  $S_C(4) = 0.97$ ). We identified corresponding firing patterns with similar properties in both the anesthetized and awake datasets (**Figure 3.6d-h**). Although mean firing rates were considerably higher in the awake state (**Figure B.3a**, **Figure 3.6d**), the direct correspondence of firing pattern types between the datasets indicates that these firing patterns are consistent across both states, and not exclusive to one.

Type-1 firing patterns exhibited low mean spontaneous firing rate (anesthetized:  $2.07 \pm 2.29$ ; awake:  $7.15 \pm 6.78$ ; **Figure 3.6d**), irregularity, indicated by a high coefficient of variation (anesthetized:  $1.53 \pm 0.58$ ; awake:  $1.57 \pm 0.57$ ; **Figure 3.6e**),

### 3. RESULTS

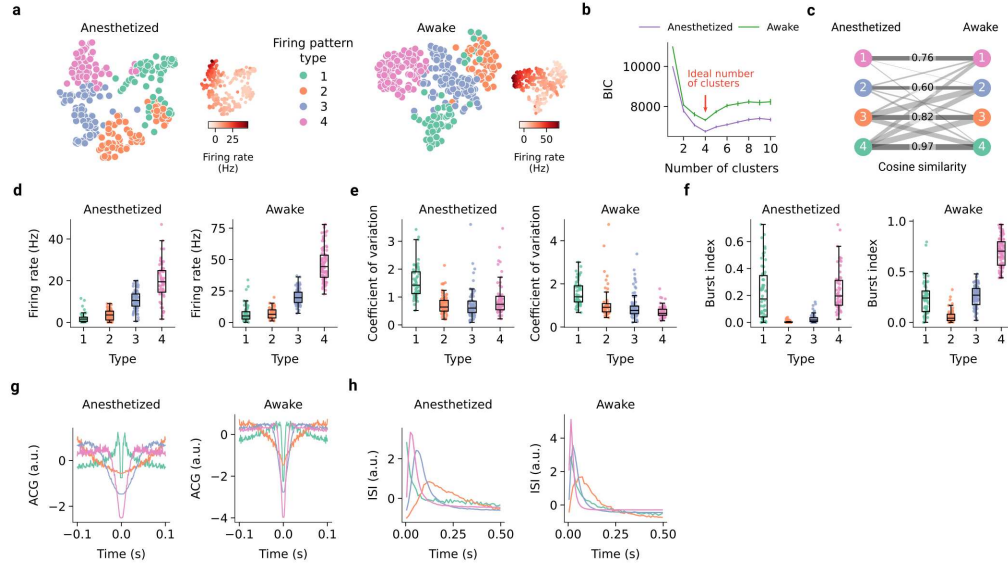


**Figure 3.6: GMM clustering on spontaneous firing features reveals four LHB neuron firing pattern types.** (a) t-SNE embedding of spontaneous firing patterns of lateral habenula (LHB) neurons recorded under anesthesia (left;  $n_{\text{neurons}} = 270$ ) and awake (right;  $n_{\text{neurons}} = 295$ ). Each point represents a recording. Firing pattern types (color-coded) were determined using Gaussian Mixture Modelling (GMM) based on firing pattern features. Matching colors indicate the corresponding clusters across both conditions (see B.2.4). The insets show t-SNE embedding colored by the mean spontaneous firing rate of each neuron. (b) Bayesian information criterion (BIC) computed for different number of clusters. Arrow indicates minimum BIC and the ideal number of clusters. For each number of clusters the GMM was initialized randomly 100 times and the error bars show the confidence intervals across runs. (c) Matching cluster using the Hungarian algorithm. Computed cosine similarities of mean feature vectors from matched clusters are indicated on the lines. The line widths are relative to the extent of similarity. (d-f) Mean firing rate (d), coefficient of variation (e), and burst index (f) for each firing pattern type under anesthesia (left) and awake (right). N and conventions as in (a). (g,h) Mean autocorrelograms (g) and interspike interval distributions (h) for LHB neurons recorded under anesthesia (left) and awake (right). N and conventions as in (a).

and was bursty, as shown by a high burst index (anesthetized:  $0.22 \pm 0.20$ ; awake:  $0.24 \pm 0.17$ ; **Figure 3.6f**) and a peak at 5 ms in the ACG and ISI distributions (**Figure 3.6g,h**). Type-2 firing patterns exhibited a low mean spontaneous FR (anesthetized:  $3.45 \pm 2.47$  Hz; awake:  $7.37 \pm 4.23$  Hz; **Figure 3.6d**), a very low burst index (anesthetized:  $0.00 \pm 0.01$ ; awake:  $0.06 \pm 0.07$ ; **Figure 3.6f**), and a tendency to fire regularly, as indicated by the ACG and ISI distributions (**Figure 3.6f-h, h**). Type-3 firing patterns showed a higher mean firing rate (anesthetized:  $10.67 \pm 4.61$ ; awake:  $20.18 \pm 5.80$ ; **Figure 3.6d**), a moderate burst index (anesthetized:  $0.03 \pm 0.03$ ; awake:  $0.26 \pm 0.10$ ; **Figure 3.6f**), and a low coefficient of variation (anesthetized:  $0.74 \pm 0.54$ ; awake:  $0.93 \pm 0.54$ ; **Figure 3.6e**). Type-4 firing patterns displayed a high firing rate (anesthetized:  $19.57 \pm 9.25$ ; awake:  $46.42 \pm 13.32$ ; **Figure 3.6d**), a high burst index (anesthetized:  $0.24 \pm 0.16$ ; awake:  $0.69 \pm 0.15$ ; **Figure 3.6f**), a low coefficient of variation (anesthetized:  $0.95 \pm 0.62$ ; awake:  $0.69 \pm 0.24$ ; **Figure 3.6e**), and an ISI distribution peaking at 25 ms (**Figure 3.6h**).

### 3.2.2 Firing pattern Type-1 LHb neurons are preferentially modulated by noxious stimuli

LHb neurons are known to exhibit varied excitatory and inhibitory responses to noxious sensory stimuli, such as foot shocks ([83, 215]; example responses shown in **Figure 3.7a**). To investigate how different spontaneous firing pattern types relate to the response to noxious stimuli, we computed a FS modulation index (FSMI; see methods B.2). In a previous study, we demonstrated that the FSMI is negatively correlated with the spontaneous mean firing rate of LHb neurons [84]. Since neurons with different spontaneous firing patterns vary in their mean firing rates, we expected to observe differences also in their FS responses. Indeed, we observed varied responses across all four spontaneous firing patterns, with each

### 3. RESULTS

---

type containing both FS-excited ( $\text{FSMI} > 0$ ) and inhibited ( $\text{FSMI} < 0$ ) neurons. For all firing pattern types, mean FS responses were dominated by short-latency excitatory components (**Figure 3.7b,c**). Interestingly, neurons with spontaneous firing pattern Type-1 exhibited, on average, FS responses with a prominent long-latency excitatory component (peaking at 210 ms; **Figure 3.7b,c**). Further comparison of Type-1 against other spontaneous firing patterns revealed a higher FS modulation index (Type 2: mean difference = -0.205, 95% CI [-0.318, -0.091],  $p < 0.001$ ; Type-3: mean difference = -0.313, 95% CI [-0.423 -0.203],  $p < 0.001$ ; Type-4: mean difference = -0.262, 95% CI [-0.379 -0.144],  $p < 0.001$ ; **Figure 3.7d,e**). These results indicate that Type-1 neurons are comparatively more strongly and sustainedly modulated by FS stimulation. Further, FS modulation was positively correlated with burstiness (Pearson  $r = 0.29$ ,  $p = 7.94 \times 10^{-11}$ ) and negatively correlated with firing rate (Pearson  $r = -0.36$ ,  $p = 4.47 \times 10^{-7}$ ), both prominent in Type-1 neurons (**Figure B.2e**).

#### 3.2.3 Type-1 firing pattern LHb neurons are preferentially modulated by spontaneous behavioral state

Recent work in awake head-fixed mice has demonstrated that spontaneous behavioral state fluctuations are accompanied by brain-wide modulations of neuronal activity [385]. Notably, while this activity modulations appears to be a brain-wide phenomenon, within each brain area, only a fraction of neurons are significantly modulated [42] pointing to cell-type specific mechanisms. To investigate whether and how the different LHb neuronal types are modulated by behavioral state, we acquired juxtacellular recordings along with orofacial videography in awake head-fixed mice. Using whisker-pad motion energy to index behavioral states [42](ref Carandini Harris; Blanco-Hernandez et al., 2024), we categorized the animal's behavior into two states, 'active' and 'inactive' immobility (see methods B.2.3;

### 3. RESULTS

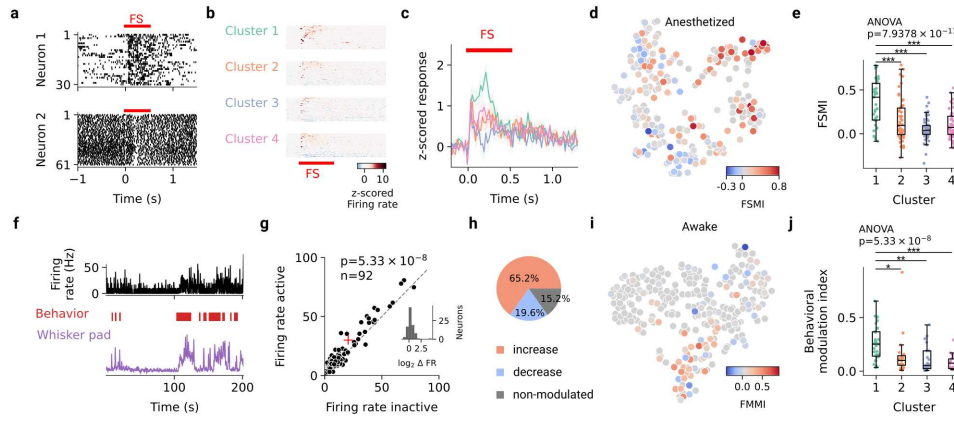


**Figure 3.7: Sensory and behavioral modulation of LHB neurons with different firing spontaneous pattern types.** (a) Example responses to foot shock (FS) stimulation for two example neurons in a spike raster plot. Onset and duration of FS stimulation indicated by the red line. (b) FS responses per cluster. Onset and duration of FS stimulation indicated by red line at the bottom. (c) Mean FS responses of LHB neurons with different firing pattern types. Red line indicates the onset and duration of FS stimulation. (d) FS modulation index (FSMI) of LHB neurons with different firing pattern types (ANOVA,  $n_{\text{neurons}} = 219$ ,  $p = 7.93 \times 10^{-11}$ ). Firing pattern Type-1 neurons were more strongly modulated in comparison to other types (Type-1 vs. Type-2: mean difference = -0.2049, 95% CI [-0.3183, -0.0914],  $p < 0.001$ ; Type-1 vs. Type-3: mean difference = -0.3131, 95% CI [-0.4232, -0.2030],  $p < 0.001$ ; Type-1 vs. Type-4: mean difference = -0.2615, 95% CI [-0.3788, -0.1441],  $p < 0.001$ ; Type-2 vs. Type-3: mean difference = -0.1082, 95% CI [-0.2044, -0.0120],  $p = 0.0206$ ). (e) Relationship of FS modulation index and burst index. Type-1 neurons show the highest modulation, while still being bursty. Pearson correlation coefficient, p-value and confidence intervals are indicated. (f) Representative recording of a LHB neuron along with facial motion. Top: firing rate; bottom: facial motion over time. Red marks indicate epochs of active states. (g) Effect of spontaneous behavioral state on LHB firing rate.  $p$  value denotes the result of the paired Wilcoxon signed-rank test,  $n = 92$ , 20.58 Hz vs. 29.97 Hz, indicated with red cross. *Inset*: Histogram of firing rate fold-change relative to inactive state ( $\Delta \text{FR} \log_2$ -ratio). (h) Percentages of neurons that significantly increase and decrease their firing rates in response to behavior onset and neurons that are non-modulated by behavioral state. (i) t-SNE embedding of spontaneous firing patterns color coded by facial motion modulation index (FMMI;  $n_{\text{neurons}} = 92$ ; neurons without facial motion recorded depicted in gray). (j) Facial motion modulation index of LHB neurons with different firing pattern types (ANOVA,  $n_{\text{neurons}} = 92/295$ ,  $p = 1.55 \times 10^{-4}$ ). Firing pattern Type-1 neurons were more strongly modulated in comparison to other types. (Type-2: mean difference = -0.118, 95% CI [-0.229, -0.008],  $p = 0.031$ ; Type-3: mean difference = -0.151, 95% CI [-0.259, -0.042],  $p = .003$ ; Type-4: mean difference = -0.180, 95% CI [-0.296, -0.063],  $p < 0.001$ ).

### 3. RESULTS

---

**Figure 3.7f**). On average, the onset of facial motion was associated with a significant net increase in LHb activity (Wilcoxon test,  $n_{\text{neurons}} = 92$ ,  $p = 5.22 \times 10^{-8}$ , 20.58 vs. 29.97 Hz) (**Figure 3.7g**). However, individual LHb neurons were diversely modulated by behavioral states: the majority of neurons (60/92) showed a significant increased in mean firing rates during periods of active immobility, while a minority exhibited reduced mean firing rates (18/92), or no significant modulation (14/92; **Figure 3.7h**; see Methods). Notably, these findings are consistent with LHb responses to FS, where most neurons were excited, while a minority were inhibited or non-responsive. [83, 84]. To quantify the degree of behavioral state modulation, we computed for each neuron a behavioral modulation index (FMMI; see methods B.2.3). We found that compared to the other clusters, Type-1 neurons were more strongly modulated by spontaneous behavior, as indicated by a comparatively higher FMMI (ANOVA,  $p < 0.005$ ,  $n_{\text{neurons}} = 92$ ; Type-2: mean difference=-0.118, 95% CI [-0.229,-0.008],  $p = 0.0306$ ; Type-3: mean difference=-0.151, 95% CI [-0.259,-0.042],  $p = 0.0027$ ; Type-4: mean difference=-0.180, 95% CI [-0.296,-0.063],  $p = 0.6 \times 10^{-3}$ ; **Figure 3.7i,j** with results of Tukey's HSD). These results indicate that LHb neurons are differentially modulated by spontaneous behavioral state fluctuations, with Type-1 neurons exhibiting, on average, the strongest modulation with the FMMI.

### Discussion

In the present study, we investigated the electrophysiological diversity of neurons in the LHb, a brain region involved in processing aversive stimuli and regulating motivated behaviors. Previous work has revealed a high degree of morphological, electrophysiological and molecular diversity within LHb circuits. Whether this diversity also applies to in-vivo firing patterns had remained largely unexplored. By applying an unsupervised clustering approach to in-vivo electrophys-

iological data, we identified four distinct spontaneous firing pattern types in LHb neurons. Notably, these firing pattern types were consistently observed in both anesthetized and awake animals (**Figure 3.6**), indicating that this diversity might be imposed by structural determinants. Indeed, our observations are consistent with previous ex-vivo studies, which identified four distinct firing pattern types in the rodent LHb [270, 430], and transcriptomic studies, which also identified four molecularly-defined LHb cell types, each with its own connectivity [422]. We note, however, that despite the above correspondences, our observations are based on unsupervised clustering of firing patterns. Hence, whether the four clusters correspond to distinct LHb cell types, or rather map onto a continuum of neuronal features, remains to be investigated. Moreover, whether the cluster affiliation of the different LHb neurons is static (i.e. pre-determined by e.g. neuronal intrinsic properties and/or synaptic inputs) or dynamically modulated (e.g. by learning and behavior [84]), remains to be investigated. We speculate that the diversity in firing patterns observed in-vivo might correlate to the known diversity of intrinsic neuronal properties, morphologies and input connectivity described ex-vivo [195, 429, 430]. Indeed, the tendency of neurons to fire spike bursts [171] and input-output-connectivity [194] are anatomically organized within the LHb [172]. Future work will be required for testing this hypothesis, and for resolving possible structure-function relationships within LHb circuits.

Spontaneous fluctuations in behavioral state modulate neuronal activity throughout the brain, yet only a small fraction of neurons in each brain structure are strongly modulated [385]. Indeed, in line with previous work, we found that only a minority of neurons within the LHb was particularly responsive to aversive stimuli in anesthetized mice, and behaviorally modulated in awake mice. Interestingly, under both anesthetized and awake conditions, these modulated neurons were preferentially recruited from Type-1 firing patterns (**Figure 3.7**). Compared

### 3. RESULTS

---

to the other clusters, these neurons displayed a higher tendency to fire spike bursts (**Figure 3.6**). Since spike bursts have been linked to plasticity [84] and excitability [202], our data indicate that this minority Type-1 LHb neurons might play a fundamental role in information encoding within the LHb. This hypothesis is indeed consistent with a large body of previous work, which demonstrated that minorities of excitable and plastic cells dominate information encoding and processing in different brain areas like neocortex [449], hippocampus [457], and amygdala [122]. Neuronal spike bursts are also particularly efficient at overcoming synaptic failures and transmitting information to downstream targets [204]. Indeed, bursts in LHb neurons efficiently inhibit reward-related downstream targets, and have been shown to play an important role in depression-like behavioral states [447]. Hence, Type-1 neurons might also play a dominant role in pathophysiological LHb states. We speculate that modulating the activity of Type-1 neurons may be a particularly powerful approach for modulating LHb computations - as demonstrated for example, in the hippocampus, neocortex and amygdala [9, 122, 143], where artificial modulation of ‘active minorities’ has been shown to be sufficient for modifying memory content and for driving behaviour. Whether Type-1 LHb neurons can be preferentially tagged with immediate early gene expression [231], as in cortico-hippocampal and amygdalar networks [398] remains to be investigated.

In summary, our data provide insights into the diversity of LHb in-vivo firing patterns, and point to low-firing, bursty (Type-1) neurons playing a unique role in information encoding within LHb circuits.

### 3.3 Study III: Plasticity of neuronal dynamics in the lateral habenula for cue-punishment associative learning

Published as: Congiu, M., Mondoloni, S., Zouridis, I.S., **Schmors, L.**, Lecca, S., Lalive, A.L., Ginggen, K., Deng, F., Berens, P., Paolicelli, R.C., Li, Y., Burgalossi, A., & Mameli, M. (2023). Molecular Psychiatry 28, 5118–5127. <https://doi.org/10.1038/s41380-023-02155-3>

#### Author contributions

Contributions according to the CRediT taxonomy [8]:

**Conceptualization:** M.C., M.M.; **Data Curation:** **L.S.**, I.S.Z., A.B.; **Formal Analysis:** **L.S.** (data preprocessing, neural population analysis), I.S.Z. (data preprocessing, single cell activity analysis); **Funding Acquisition:** P.B., A.B., M.M.; **Investigation:** M.C. (electrophysiological, behavioural, optogenetic manipulations, two-photon imaging), S.M. & S.L. & A.L.L. & K.G. (complementary experiments), I.S.Z. (complementary experiments, data analysis for electrophysiological data), F.D. (generation of the 5HT biosensor); **Methodology:** M.C., M.M., P.B. & R.C.P. & Y.L. (know-how and tools for analyses and imaging), A.B. (support for electrophysiological recordings and their analysis); **Project Administration:** M.C., M.M., P.B. & R.C.P. & Y.L. (know-how and tools for analyses and imaging), A.B. (support for electrophysiological recordings and their analysis); **Resources:** P.B., R.C.P., Y.L.; **Software:** M.C., **L.S.**, I.S.Z.; **Supervision:** P.B., R.C.P., Y.L., A.B., M.M.; **Validation:** **L.S.**, I.S.Z., A.B.; **Visualization:** M.C., M.M.; **Writing – Original Draft:** M.C., M.M.; **Writing – Review & Editing:** all authors.

## Introduction

Pavlovian conditioning represents a temporally trackable brain function in which sensory cues are associated with incentive stimuli enabling individuals to anticipate upcoming threats and rewards. This process is at the basis of complex behavioural outcomes including defensive responses such as freezing or avoidance [82, 158, 217, 382]. During Pavlovian conditioning, a sensory cue, the conditioned stimulus (CS+), is associated with an unconditioned stimulus (US), an air-puff directed to the eye for instance [382]. Upon re-exposure to the CS+ only, an eyeblink is elicited in the prediction of the subsequent airpuff, indicative of associative learning and the establishment of anticipatory behaviour [82, 256, 382]. Past work found lateral habenula (LHb) neurons with potentiated synaptic glutamate transmission after learning along with CS-mediated phasic excitation of the LHb [212, 410]. Notably, this punishment-predictive cue-mediated LHb neuronal excitation is conserved across species as it is present in humans, non-human primates, rodents, and fish [12, 211, 256, 424]. The LHb receives negativere-lated information from a subcortical network including hypothalamic, limbic, and basal ganglia nuclei [213]. The neural activity within the LHb represents punishments and negative affective states at both sub-second and slower timescales [172]. While LHb neurons phasically increase their activity in response to punishments, persistent synaptic adaptations and enhanced neuronal activity emerge during elevated stress conditions [90, 224, 256, 289, 459]. Altogether, these observations support a causal link between the excitation of LHb and the encoding of negative valence and affect [172]. Recent advances pinpoint, however, a degree of diversity in the LHb that is based on both molecular and functional features [125, 150, 394, 422]. This suggests that excitation across all LHb cells seems unlikely to be the sole mechanism to support cue-punishment associative learning.

Independent and differentially adaptive neuronal populations allow learning and memory storage [144], yet whether this applies to habenula and through which mechanisms remain unknown. Here, we combined temporally- and spatially-controlled analysis of both neuronal and neurotransmitter dynamics together with an aversive Pavlovian conditioning task in awake rodents to study LHb contribution to learning. We observed that punishments and punishment-predictive cues recruit both pre- and postsynaptic adaptations that enable discrete cell ensembles in the LHb to encode punishment-related associative memories.

## Results

### 3.3.1 Punishment drives distinct LHb neuronal populations with excitation or inhibition

Punishments generate excitation of LHb neurons as measured by recordings of neuronal activity in anesthetized mice and analysis of bulk calcium ( $\text{Ca}^{2+}$ ) signals through fiber photometry [215, 410, 424]. This led to a general framework whereby LHb encodes aversion through its excitation. However, recent work indicates that the response properties of individual LHb neurons to aversive stimuli might be more heterogeneous [34, 83, 109]. To systematically explore this possibility, we performed juxtacellular recordings from single LHb neurons in anesthetized mice, while delivering foot shock stimuli (Figure 3.8a). In line with previous observations [83, 225], we found that foot shock stimuli led to a rapid increase in spiking activity in a majority of LHb neurons (Figure 3.8b). Notably, these excitatory responses were heterogeneous in their kinetics likely due to diverse ionic conductances or synaptic input organization onto LHb cells (Figure 3.9a-f). Aside from the foot shock-excited population, a distinct subset of LHb cells was instead significantly inhibited by the stimuli (Figure 3.8c,d). To explore

### 3. RESULTS

---

whether foot shock responses relate to in-vivo firing patterns of LHb neurons, we used Stochastic Neighbor Embedding [161] and visualized the neurons in two-dimensional space (Figure 3.8e; see Methods). Mapping of foot shock responses onto the tSNE embedding, which was created based on spontaneous spiking features alone (see Methods), revealed a gradual organization of foot shock responses that correlated with spontaneous firing rates (Figure 3.8e). Indeed, foot shock modulation and spontaneous firing rates were negatively correlated (Figure 3.8f; [83]). Thus, these data indicate that in the LHb, punishments generate rapid and opposing responses following a non-random organization where foot shock excitation is more common in low-firing neurons while inhibition in high-firing cells. Altogether, this diversity of responses expands the current predominant framework in which the LHb encodes aversion solely through neuronal excitation.

#### 3.3.2 Disrupting LHb function affects cue discrimination during punishment associative learning

To understand whether the functionally diverse signatures of LHb activity in relation to punishments hold true in awake and behaving mice, we modified a classical conditioning task designed for investigating value coding in dopamine neurons with the objective of studying neuronal dynamics at a single-cell resolution [82, 382]. Head-fixed mice were trained to associate one conditioned auditory stimulus (CS+), but not another (CS-), with an airpuff directed to the eye (Figure 3.10a,b). Following two sessions of habituation to the auditory cues and two conditioning sessions where only the CS+ was contingent to airpuff presentation, mice displayed anticipatory eyeblink to the CS+ only, indicative of CS-US association (Figure 3.10c; Supplementary Figure C.1a-d). To quantify congruous learning, we tracked and analyzed the eye area during CS to compute the discrimination score (CS- average – CS+ average), a measure significantly higher

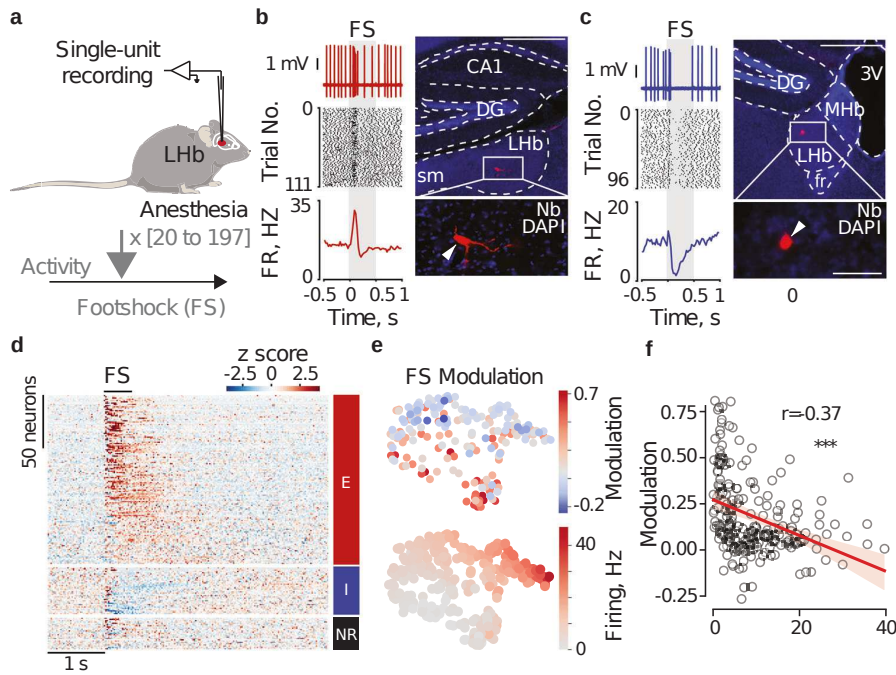


**Figure 3.8: Lhb neurons can show excitatory or inhibitory responses to foot shock.** (a) Experimental protocol related to the extracellular single-unit electrophysiological recordings in anesthetized mice; for each recorded neuron, mice received between 20 and 120 contralateral foot shocks. (b) Electrophysiological trace, raster plot, average firing response, and picture of an example foot shock-excited neuron labeled in the Lhb (scale bar top=500  $\mu\text{m}$ , bottom=100  $\mu\text{m}$ ). (c) Same as b but for a foot shock-inhibited neuron. (d) Heatmap of single neuron z-scored average response to FS presentation ( $n_{\text{neurons}}=196$ ;  $n_{\text{neurons EXC}} = 134$ ,  $n_{\text{neurons INH}} = 37$ ,  $n_{\text{neurons NR}} = 25$ ). (e) t-SNE plots with superimposed color-coded FS-modulation index (top) and baseline firing rate (bottom). (f) scatter plot with linear correlation of firing rate and modulation index with all cells ( $n_{\text{neurons}} = 196$ ,  $***p < 0.001$ ).

### 3. RESULTS

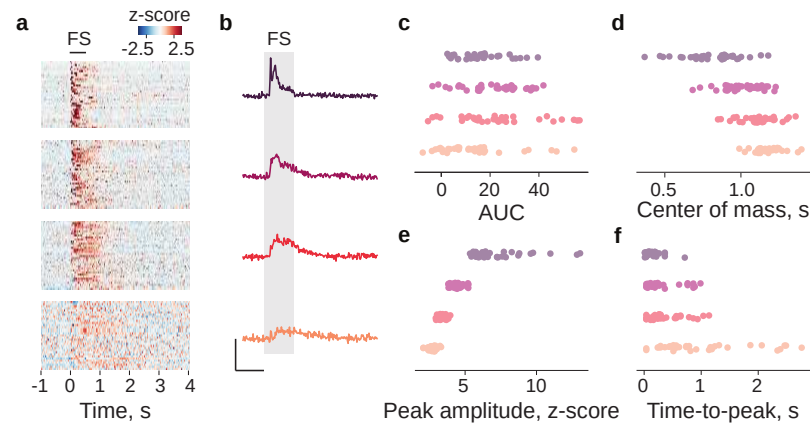


Figure 3.9: **Kinetics heterogeneity of FS-excitatory responses in LHb neurons.** (a) Z-scored peristimulus time histogram for all FS-excited neurons in the LHb (same as in Figure 3.8d) divided in four equally-sized groups. Same n and conventions as in Figure 3.8d. (b) Mean peristimulus time histograms for the four groups indicated in (b). (c-f) Quantification of FS-response properties for the four groups indicated in (a, b). ‘Area under the curve’ calculated as the summed z-scored firing rate for every trial within the 0-3 s window following FS onset; ‘Center of Mass’ calculated as the area under the curve weighted by the time vector and divided by the number of bins; ‘Peak Amplitude’ represents the normalized maximal response amplitude within the 0-3 s window following FS onset; ‘Time-to-Peak’ represents the time from stimulus onset to the response maximum. Same color code as in (b).

after conditioning compared to habituation (Figure 3.10d; Supplementary Figure C.1e,f). We next aimed to probe the necessity of LHb for cue discrimination during associative learning by optically reducing its activity. We transduced LHb neurons with a light-driven chloride pump (orange-red spectrum of activation) via the infusion of rAAV8-hSyn1-Jaws-EGFP (Figure 3.10e). Four weeks later, we assessed basal LHb neuronal activity using single-unit recordings in anesthetized mice [215]. Light at 638 nm produced a time-locked reduction in the basal firing rate of LHb neurons (Supplementary Figure C.2a-d). Thus, Jaws efficiently reduces LHb neuronal activity. Next, we chronically implanted Jaws-expressing mice with a single fiber optic above the LHb (Figure 3.10e; Supplementary Figure C.2e,f). Shining light at 638 nm to inhibit LHb function throughout CS+ and US presentation diminished cue discrimination during conditioning (Figure 3.10f; Supplementary Figure C.2g). We acknowledge minimal Jaws spread in the paraventricular nucleus of the thalamus (PVT), a structure as well relevant for cue-punishment association [305]. Although we cannot completely rule out PVT con-

tribution in our task, the presented experimental evidence, the depth of red-light penetration, and the location of fiber probes support the LHb as a player in establishing behavioral cue discrimination emerging during associative learning.

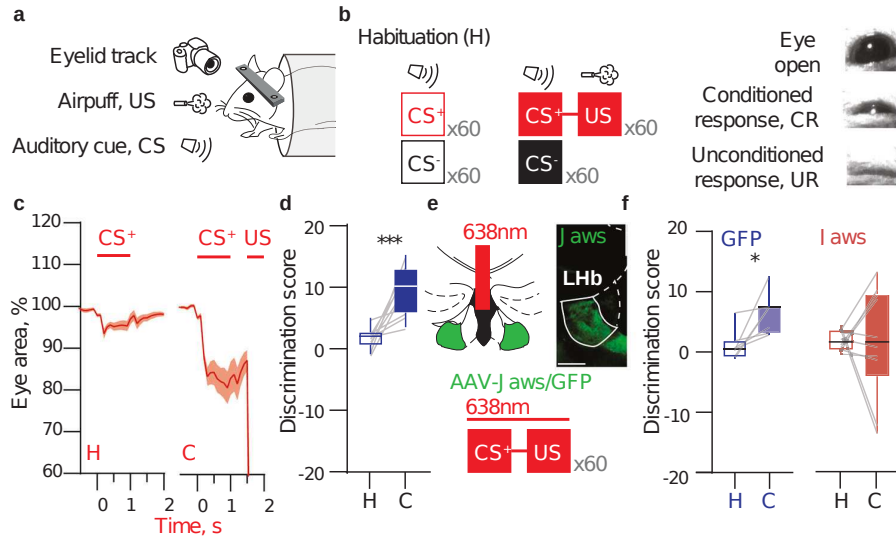


**Figure 3.10: LHb integrity is necessary for cue discrimination during Pavlovian conditioning.** (a) Experimental protocol of the behavioral setup for the Pavlovian conditioning task in awake head-restrained mice. (b) Left, schematic representing the structure of conditioning; right, example pictures of the behavioral responses during the task (c) Time course of the average eye area changes across CS+ trials during habituation (H) and conditioning (C) (n<sub>mice</sub>=10). (d) Box-plot of the discrimination score during habituation and conditioning (n<sub>mice</sub>=10; paired t test,  $t_9 = 4.92$ ,  $***p < 0.001$ ). (e) Top, schematic of optogenetic inhibition and representative coronal section of Jaws expression in the LHb and fiber track (scale bar: 200  $\mu$ m); bottom, schematic of the experimental protocol for optogenetic LHb inhibition. (f) Box-plots of the discrimination score for GFP mice (n<sub>mice</sub>=7; paired t test,  $t_6 = 3.20$ ,  $*p = 0.019$ ) and Jaws mice (n<sub>mice</sub>=10; paired t test,  $t_9 = 0.059$ ,  $p = 0.954$ ).

### 3.3.3 LHb neuronal populations develop opposing responses to punishment-predictive cues across learning

To track LHb neuronal activity in behaving mice throughout Pavlovian conditioning, we injected a viral construct encoding the Ca<sup>2+</sup> indicator GCaMP6f (rAAVdj-GCaMP6f) into the LHb [410]. Next, we implanted a microendoscopic lens allowing chronic optical access to the LHb through a two-photon microscope equipped with a 20X air objective (Figure 3.11a-c). Post hoc analysis reflected that the majority of the LHb neuronal population expressing GCaMP6f was glutamatergic

### 3. RESULTS

---

and not GABAergic (i.e. EAAC1+ and GAD67-; **Figure 3.11a**; Supplementary **Figure C.3a-c**; [410]). We recorded from GCaMP6f-expressing LHb neurons across multiple fields of view (FOVs) within each mouse ( $n = 12$  mice, 21 FOVs, 339 neurons; **Figure 3.11d**). We computed cue-driven neuronal responses and found that, while a population of LHb neurons displayed a substantial increase in their fluorescence, a different neuronal ensemble exhibited a reduction in fluorescence after the presentation of the punishment-predictive cue during conditioning, compared to the habituation period (**Figure 3.11d**). Notably, CS + -driven neuronal responses before conditioning were likely due to the saliency of the stimulus and not an intrinsic aversive component of the auditory cue. Indeed, mice did not show avoidance in a real-time place paradigm test (Supplementary **Figure C.4a**). Notably, CS+ responsive neurons were also tuned in the same direction by the air puff (Supplementary **Figure C.4b,c**). Tracking and quantification of fluorescence signals during the presentation of CS+ and CS- were indicative of plasticity after conditioning (**Figure 3.11e**). Indeed, we found that throughout habituation to conditioning, 121/339 neurons displayed an increase in Ca<sup>2+</sup>-mediated fluorescence (likely corresponding to higher excitation) time-locked to the CS+, whereas 47/339 showed a significant reduction in fluorescence to the CS+ across learning (likely corresponding to larger inhibition; **Figure 3.11f,g**;  $p \leq 0.05$ , CS-evoked signals versus baseline, rank-sum test). Thus, distinct populations of LHb neurons developed an enhancement of existing excitatory or inhibitory responses as well as new excitatory or inhibitory responses to a punishment-predictive cue across learning. Different functional responses in the LHb may anchor on territorial distribution, which embeds LHb neurons within independent circuit connectivity loops [172]. To examine the anatomical localization of excited and inhibited neurons, we matched the fields of view of GCaMP6f-expressing neurons with the respective Gradient Index (GRIN) lens edges and the anatomical

borders of LHb (See methods). CS+-activated neurons undergoing potentiation after learning largely overlaid the lateral territory of the LHb. In contrast, CS+-inhibited neurons after conditioning were rather medially-located (Supplementary [Figure C.4d,e](#)). Notably, when preparing acute brain slices and patching neurons located in the lateral aspect of the LHb after CS- airpuff association, we observed increased AMPA/GABA and AMPA/NMDA ratios along with larger amplitudes of spontaneous excitatory currents (sEPSCs) compared to control mice (CS and airpuff non-contingent; Supplementary [Figure C.5a-e](#)). In contrast, the paired-pulse ratio of evoked AMPA current and the frequency of sEPSCs remained unaltered (Supplementary [Figure C.5d,e](#)). Altogether, distinct LHb neurons undergo a CS+-driven increased excitation or inhibition, are anatomically segregated, and develop postsynaptic adaptations across punishment associative learning. We then predicted that cue responses developing through learning may represent information related to eye blinking and, therefore, cue discrimination (CS+ versus CS-). Using the CS- driven responses of each LHb neuron within each trial, we trained a support vector machine (SVM) decoder to predict, on a trial-by-trial basis, whether or not mice displayed a conditioned response. We found that activity dynamics of subpopulations of excited and inhibited LHb neurons predicted whether animals' eye-blinked during the cue as compared to the shuffled data from the same neuron ([Figure 3.11h](#)). In line with the observation that optically perturbing LHb function disrupts cue discrimination, these findings suggest that cue responses in LHb neurons represent information related to the discrimination of independent cues, and therefore proper cue-punishment association.

### 3. RESULTS

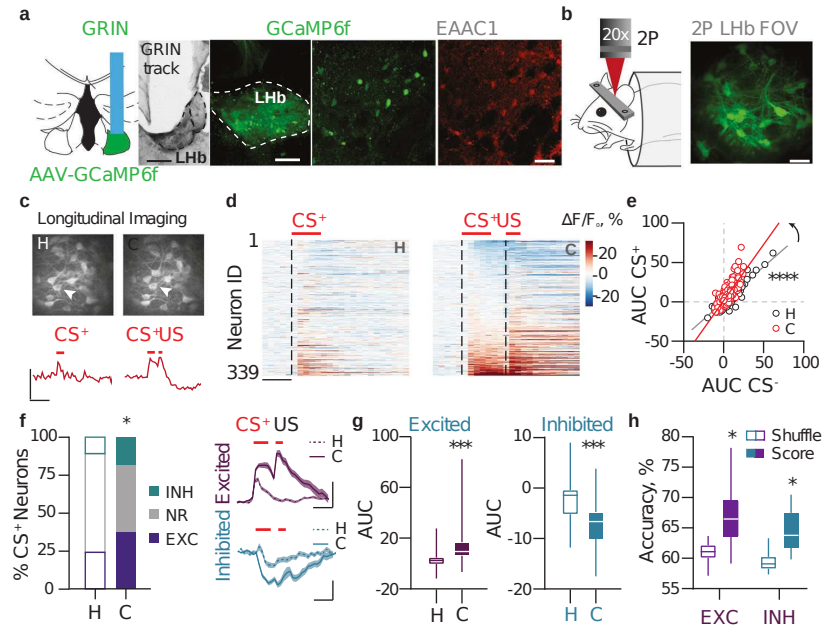


**Figure 3.11: Single neuron inhibitory and excitatory responses adapt during associative learning.** (a) Schematic of the experimental preparation for the LHB two-photon calcium imaging, example coronal histological section of GRIN lens positioning above LHB (scale bar: 200  $\mu\text{m}$ ), GCaMP6f expression on LHB neurons (scale bar: 100  $\mu\text{m}$ ) and staining of the EAAC1 glutamatergic marker (scale bar: 30  $\mu\text{m}$ ). (b) Schematic representation of the experimental setting for two-photon calcium imaging recordings in head-restrained awake mice (left) and standard deviation projection of an example field of view (FOV) (right). (c) Example of longitudinal tracking of the same FOV across habituation and conditioning sessions (top) and average response of the indicated neuron (arrowhead; bottom; scale bar: 10  $\Delta F/F_0$  %, 3 s). (d) Heatmaps of the average  $\Delta F/F_0$  response of single neurons during habituation (left) and conditioning (right) (scale bar: 1 s). (e) Scatter plot of the area under the curve (AUC) during CS+ and CS- for habituation (white dots) and conditioning (red dots) ( $F(1,674) = 86.76$ ,  $*p < 0.0001$ ,  $n_{\text{mice}} = 12$ ,  $n_{\text{neurons}} = 339$ ). (f) Left, contingency table of the proportion of neurons statistically inhibited, not responding or excited during CS+ presentation ( $\chi^2 = 8.10$ ,  $*p = 0.018$ ); right, time-course of the grand-average responses to CS+ trials during habituation and conditioning trials for the excited (top,  $n_{\text{neurons}} = 118$ ) and inhibited neurons (bottom,  $n_{\text{neurons}} = 45$ ). (g) Box-plot of the population average AUC response during CS+ presentation for the excited (left; paired t-test,  $t_{117} = 9.08$ ,  $***p < 0.0001$ ) and inhibited neurons (right; paired t-test,  $t_{44} = 6.70$ ,  $***p < 0.0001$ ). (h) Box-plot of the accuracy score for the excited and inhibited neurons ( $n_{\text{cells}}$ : excited=47; inhibited=10; Two Way ANOVA RM and Sidak's multiple comparison test, shuffle vs score  $F(1,55) = 93.84$ ,  $*p < 0.0001$ ).

### **3.3.4 Acetylcholine, but not glutamate, GABA and serotonin, release in LHb mirrors neuronal dynamics after conditioning**

Synaptic adaptations within the LHb after learning rely on postsynaptic mechanisms, including AMPA or GABA receptor trafficking/efficiency rather than changes in glutamate or GABA release [208, 410]. On the other hand, neuromodulators including serotonin and acetylcholine (ACh), are capable of modulating synaptic function and contributing to motivated behaviours [354, 355, 441]. Thus, it is plausible that during cue-punishment Pavlovian conditioning, while glutamate and GABA release remain stable, neuromodulation would instead adapt and match the postsynaptic responses, potentially representing a gating mechanism for plasticity and learning. To test this possibility, we leveraged recent advances in the readout of fluorescence intensity-based genetically encoded sensors with fiber photometry in head-fixed mice during the Pavlovian discrimination task [225]. The genetically encoded sensor iGluSnFR enables glutamate release detection, the iGABASnFR2 detects GABA release instead, while GRAB5-HT2h and GRABACH3.0 permit measurement of serotonin and ACh, respectively [183, 253, 254, 423, 451]. We expressed each of the sensors in an independent group of mice through viral injections in the LHb (**Figure 3.12a**). Presentation of an unpredicted airpuff or a neutral auditory cue produced rapid glutamate, GABA, serotonin, and ACh transients (Supplementary **Figure C.6a-d**). During conditioning, iGluSnFR, iGABASnFR2, and GRAB5-HT2h-mediated fluorescence transients remained comparable between CS+ and CS- (**Figure 3.12b-e** and Supplementary **Figure C.6a-c**). Instead, similarly to the postsynaptic GCaMP6f responses, ACh transients significantly increased after learning in response to CS+ compared to CS- (**Figure 3.12f** and Supplementary **Figure C.6d**). This indicates

### 3. RESULTS

that postsynaptic adaptations of CS+-driven excitatory and inhibitory responses in the LHb during Pavlovian conditioning occur along with a presynaptic increase of cholinergic signaling, but not glutamate, GABA and serotonin transmission.

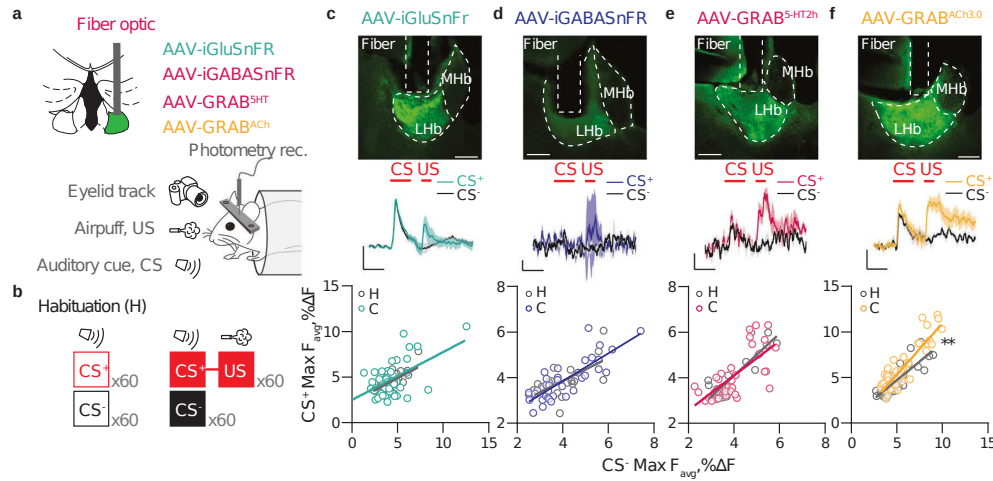


**Figure 3.12: Presynaptic ACh release in the LHb, but not glutamate, GABA, or serotonin, reflects cue discrimination.** (a) Schematic of experimental setup for the photometric recordings of neurotransmitter biosensors dynamics in the LHb. (b) Schematic of the experimental setting in head-restrained mice. (c-f) Top, histological example of the expression of the neurotransmitter sensor and fiber placement in the LHb; middle, grand-average time-course of the  $\Delta F/F_0$  response for CS+ and CS-; each trace is normalized by the maximum amplitude of the CS- response ( $n_{\text{trials}}:CS+=60$ ,  $CS-=60$ , legends =1 %, 1 s); bottom, scatter plot of the average peak  $\Delta F/F_0$  response for CS+ and CS- on a 5 trial bin for habituation and conditioning trials (C, iGluSnFR,  $n_{\text{mice}} = 5$ ,  $F(1,86) = 0.08$ ,  $p = 0.77$ ; D, iGABASnFR2,  $n_{\text{mice}} = 3$ ,  $F(1,50) = 0.39$ ,  $p = 0.53$ ; E, GRAB5-HT2h,  $n_{\text{mice}} = 3$ ,  $F(1,50) = 0.59$ ,  $p = 0.45$ ; F, GRABACH3.0,  $n_{\text{mice}} = 6$ ,  $F(1,116) = 8.13$ ,  $*p = 0.0052$ ).

### Discussion

This study reveals unprecedented and remarkable diversity in neuronal responses to punishments and their predictive cues in the LHb, unraveling complex opposing signals which may contribute to negative affect-related behaviours in health and disease. We found that LHb neurons, generally thought to be excited by aversive events, physically respond with excitatory or inhibitory signals to punishment as measured by in vivo recordings from identified LHb neurons. To extend the behavioural relevance of these findings, we show that an appropriate function of LHb is required for proper cue discrimination in a Pavlovian task where mice learn to

associate a specific cue to an upcoming punishment. LHb neuronal ensembles longitudinally recorded during learning expressed opposing cue-driven responses and adaptations with enhanced excitation and inhibition. Our data reveal that changes in excitatory and inhibitory responses in distinct LHb neuronal populations guide cue-punishment associations and consequent anticipatory behavior. These post-synaptic adaptations mirror the dynamics of specific presynaptic signaling onto LHb neurons whereby the steady glutamate and GABA release contrasts with plastic ACh signaling. These data identify novel neuronal mechanisms underlying negative valence encoding in the LHb. More broadly, our findings suggest that LHb responses to aversive stimuli, often considered as homogeneously excitatory, are instead diverse and plastic and that, along with neuromodulatory signaling in LHb, adapt throughout learning processes. Altogether, this reflects specialized and complex LHb organization and functions which likely coordinate specific behavioral outcomes (i.e. anticipatory behaviours).

The LHb responds to aversive external stimuli with a phasic increase in neuronal activity in both awake and anesthetized animals [215, 256, 424], a functional feature independent of the nature of the negative stimulus. Accordingly, foot shock, air puff, and radiant heat promote excitation of the LHb [173, 359, 410]. Thus, excitation in the LHb is defined to encode negative valence, and, if persistent, to underlie negative affect in psychiatric states [172]. In the present work, we systematically characterized neuronal responses in the LHb by using in-vivo juxtacellular recordings during foot shock stimulation in anesthetized animals. Our dataset of 200 neurons extends previous observations [83, 213] and shows that neuronal responses to foot shock are diverse, and represented by both excitation and inhibition [83]. Notably, a similar diversity of responses was also observed in awake animals [225]. After recording more than 300 LHb cells during a Pavlovian task under a two-photon microscope, the analysis recapitulates the

### 3. RESULTS

---

functional diversity and unravels how cue-driven excitation and inhibition relate to punishment-mediated transients at the single-cell level. Notably, single-cell responses (both excitatory and inhibitory) to cues were present already during the habituation session, suggesting that LHb neurons may encode the salience of a stimulus and not solely its valence. This is in line with work performed both in rodents and non-human primates [57, 225]. In the latter, phasic salience-related signals are described in the habenula-dopamine pathway – a process that motivates anticipation [57], thus overall supporting the data presented within this work.

The cue-driven excitation and inhibition emerging across the LHb both during habituation and conditioning raise the question of which determinants underlie this diversity. A plausible scenario is that excited and inhibited cells are part of distinct neuronal networks. Reconstruction of the anatomical position of each neuron recorded under the two-photon microscope unravels that inhibited cells occupy the medial territory of the LHb [83], conversely to the excited population located rather laterally. Subcortical innervation onto LHb follows a topographical map with synaptic inputs from the basal ganglia impinging specifically onto the lateral division and the bed nucleus of the stria terminalis rather limited to the medial aspect [224, 289, 355]. Similarly, medially-located neurons synapse downstream to midbrain dopamine neurons and raphe serotonin neurons, while lateral LHb cells send their axons to midbrain GABA neuronal populations [215, 249]. Thus, connectivity-based differences may underly functional signatures for learning in LHb.

In addition, or alternatively, the functional diversity may anchor on transcriptomics differences whereby independently of the neuronal circuit integration, excited and inhibited neurons may differ based on their genetic nature [150, 222, 422]. It remains yet challenging to differentially label neuronal populations expressing distinct plastic responses and discretely study their anatomical orga-

nization, function, and molecular identity. The further amelioration of  $\text{Ca}^{2+}$ -, neuronal-activity-, immediate early gene-driven labeling is required to fill this gap, at least in relation to the excited cells [218, 374, 408].

The tracking of the neuronal dynamics across the conditioning and the subsequent analysis unravels that a proportion of both inhibited and excited neurons represent cue discrimination and, thereby, CS+-driven anticipatory behaviours. This finding matches the observed loss of cue discrimination after LHb optical inhibition throughout learning. Notably, manipulating LHb function affects cue discrimination in rats undergoing a cocaine-seeking paradigm [178, 454] or in mice during an odor-outcome association task [401]. Altogether, this converges to clarify the contribution of the LHb in enabling discrimination among valued external cues. This offers a framework in which LHb aberrant activity may produce cue generalization – a behavioural feature of disorders including post-traumatic stress disorders, for instance. Behavioural cue discrimination after conditioning occurs along with i. a shift from a silent to a responsive state of a proportion of neurons that become either inhibited or excited and ii. the CS+-mediated increase of both inhibition and excitation. Thus, the CS+-driven plasticity of excited and inhibited cells after conditioning occurs following the direction of the initial CS+-mediated response during the habituation period. This process is different from other subcortical structures, including the amygdala, where adaptations in cue-driven responses after learning remain independent from the initial CS-mediated responses at baseline [144].

The plasticity occurring in the LHb after cue-punishment association may have a synaptic foundation, as suggested by the strengthening of neurotransmission onto cells located in the lateral aspects of LHb. These data are in accordance with a scenario in which learning to anticipate an aversive event requires synaptic potentiation of AMPA-mediated transmission in the LHb [212, 410]. While

### 3. RESULTS

---

plasticity at excitatory synapses emerges on putative excited neurons, whether increased inhibition during learning takes place along with adaptations of synaptic GABA<sub>A</sub> neuro-transmission remains to be established. This is plausible, however, as potentiation of synaptic inhibition in the LHb contributes to associative processing [208]. Altogether these data support the notion that postsynaptic adaptations mediate the establishment of cue discrimination and cue-punishment association. This is further corroborated by the observation that presynaptic glutamate and GABA signaling remain comparable between CS+ and CS-. This indicates that presynaptic fast neurotransmission does not change to support cue discrimination, which recapitulates previous findings obtained from recordings in acute brain slices [208, 410]. Our data lack, however, synaptic input specificity, thus presynaptic plasticity events that are synapse-specific cannot be ruled out. For instance, vesicular glutamate transporter-2 (Vglut2)-expressing hypothalamic neurons that project to the LHb undergo adaptations of cell dynamics during fear learning [212]. The differences between behavioural tasks (Pavlovian punishment task in this work versus fear conditioning) or the lack of input-specific assessment might be among the reasons explaining the absence of changes in glutamate release described here.

Synaptic plasticity requires, in many instances, gating processes often mediated by neuromodulatory signaling [3, 352]. Indeed, neuromodulation can prime synapses for the induction and expression of long-term synaptic adaptations [113, 352], whereby cholinergic, serotonergic, and dopaminergic signaling can control synaptic gain [41, 296, 316]. By using genetically encoded sensors for neurotransmitters in the LHb of behaving mice we detected phasic release of ACh and serotonin. While serotonin release represents a relevant modulatory pathway in the LHb [172, 354, 355], and its release occurs along cue and punishment presentation, this does not contribute to cue discrimination, as CS+ and CS- mediated serotonin

transients remained comparable during conditioning. Experimental evidence reporting ACh release in LHb during animal behaviour was so far absent. We report the unprecedented observation that ACh transients were detectable during cues and punishment presentation, raising a scenario whereby such signaling emerges during the encoding of salient events independently of their value. In addition, ACh signaling adapts throughout cue-punishment association, mirroring the plastic adaptations observed with GCaMP6f recordings. Thus, ACh release may gate synaptic plasticity in the LHb to support conditioning. While this remains the first attempt to describe ACh presynaptic dynamics, information on the postsynaptic machinery required to convey cholinergic signaling in the LHb already exists. LHb neurons become depolarized or hyperpolarized by activation of muscarinic cholinergic receptors. In addition, muscarinic receptor inhibits synaptic GABA and glutamate transmission [441]. Furthermore, blockade of LHb muscarinic signaling impaired operant cocaine behaviours in a Go/No-Go task [454]. Altogether, this suggests that functional cholinergic signaling through muscarinic receptors in the LHb regulate cocaine-mediated behavioural outcomes. It remains to be established which circuit connectivity underlies ACh release in the LHb, yet our findings highlight the importance of studying cholinergic modulation of synaptic efficacy and its role in behaviours associated with aversive states. In summary, our findings reveal functional diversity, plasticity events, and neuromodulatory mechanisms in the LHb that contribute to associative learning, cue-punishment association, and cue discrimination – further supporting the relevance of LHb for valence encoding. We postulate that anticipation of upcoming punishments relies on excited and inhibited LHb neuronal populations which play a synergistic role in coordinating learning. This provides insights into how the brain transforms neutral stimuli into punishment-predictive ones.

## 3.4 Study IV: Juxtacellular recordings from identified neurons in the mouse locus coeruleus

Published as: Zouridis, I., **Schmors, L.**, Fischer, K.M., Berens, P., Preston-Ferrer, P., Buraloss, A. (2024). European Journal of Neuroscience. <https://doi.org/10.1111/ejn.16368>

### Author contributions

Contributions according to the CRediT taxonomy [8]:

**Conceptualization:** I.S.Z., P.P.F., A.B.; **Data Curation:** I.S.Z. (annotate & produce metadata, preprocess data, maintain research data, maintain software code), **L.S.** (create & maintain software code); **Formal Analysis:** I.S.Z. (statistical analysis), **L.S.** (statistical analysis, feature extraction, classifier); **Funding Acquisition:** P.P.F., A.B.; **Investigation:** I.S.Z. (juxtacellular recordings, histology), K.M.F. (supported histology experiments); **Methodology:** I.S.Z. (electrophysiology and behaviour experiments, histology), **L.S.** (development of computational models); **Project Administration:** I.S.Z., A.B., **Resources:** P.B., A.B., **Software:** I.S.Z. (software development for statistical analysis), **L.S.** (software development for statistical analysis and computational model); **Supervision:** A.B. (main supervision), P.P.F. (co-supervised), P.B. (supervised computational models); **Validation:** I.S.Z., **L.S.**; **Visualization:** I.S.Z., **L.S.**, A.B.; **Writing – Original Draft:** I.S.Z., A.B.; **Writing – Review & Editing:** all authors.

## Introduction

The locus coeruleus (LC) is the principal source of noradrenergic neurotransmission in the mammalian central nervous system [391]. Noradrenergic LC neurons, through their widely divergent connections to virtually the entire central nervous system, are in a key position for regulating numerous brain functions, including attention, arousal and memory (for review see [36, 315, 338–340, 406]). LC neurons have been classically assumed to be a structurally and functionally homogeneous population; however, over the years, diverse research lines have revealed a high degree of heterogeneity and modularity within local LC microcircuits [72]. Anatomically, the LC is a complex of noradrenergic neurons located within the rostral hindbrain. This complex has been classically partitioned into a ‘nuclear core’ region, which contains the highest density of noradrenergic neurons (especially in murid rodents, see [242]), and additional neighbouring subregions, where the organization of noradrenergic cells is more scattered and intermingled with the non-noradrenergic surround [22, 50, 132, 139, 146, 182, 207, 237, 264, 365, 390]; for review, see [242]). Early anatomical studies divided LC neurons into several types based on their distinct morphological characteristics [78, 235]. More recent work indicates that LC neurons are also heterogeneous in terms of their molecular expression profiles, intrinsic electrophysiological properties and projection targets [71, 190, 228, 264, 314, 349, 386, 404, 405, 412]. However, whether and how this structural heterogeneity of LC neurons relates to *in vivo* activity has remained largely unresolved.

Extracellular recordings - the gold standard method in systems neuroscience, for review, see [63]) - have provided a massive amount of information about the *in vivo* activity of individual neurons in many areas of the mammalian central nervous system. These methods are particularly powerful when recordings can be

### 3. RESULTS

---

confidently targeted to the brain region of interest, and cell identity can be inferred from spike signals alone. In the neocortex and hippocampus, for example, reliable electrophysiological signatures of anatomical boundaries have been identified, which allow targeting even specific subfields or layers within these structures [141, 188, 269, 372]. Moreover, ‘ground-truth’ datasets from anatomically identified neurons in these structures have made it possible to develop electrophysiological criteria that enable to infer the putative cell identity of the extracellularly recorded units with a high degree of confidence [93, 102, 141, 154, 188, 198, 283, 304, 351].

The relatively small size of the LC, together with the lack of clear electrophysiological signatures of anatomical boundaries, has made it particularly difficult to infer noradrenergic cell identity from blind extracellular recordings. Even if recordings are targeted to the LC ‘nuclear core’, which is known to be almost exclusively composed of noradrenergic cells, extracellular spike signals are de facto sampled from a larger area [136, 154, 234] that extends well beyond the narrow LC ‘nuclear core’. Hence, to increase confidence in cell classification, it is necessary to first characterize the in vivo properties of unequivocally identified noradrenergic neurons and then test whether these features can be used for differentiating them from the local heterogeneous surround of non-noradrenergic cells. Elegant pharmacology-based approaches have been classically employed for inferring the putative noradrenergic identity of the recorded units [18, 243, 341, 357, 405]. More recent approaches involving high-density extracellular recordings, in combination with optogenetic tagging, have increased the efficiency of LC sampling and enabled the in vivo characterization of opto-identified LC units [51, 69, 159, 165, 177, 252, 259, 265, 392, 404, 412, 444, 445]. When cell-type specific properties are very distinct and under low levels of recurrent connectivity, optotagging methods can achieve excellent separation of the

extracellularly recorded units (as, e.g., for LC units from the medially located Barrington's nucleus [177]); however, if differences are more subtle and neural circuits more intermingled, extracellular optotagging becomes far from trivial and more critically dependent upon the performance of spike-sorting algorithms. The latter is known to be limited under conditions of high 'background' spiking activity, high cell densities and synchronous firing (as is the case in the LC) as well as during photostimulation [128, 328, 329, 380]. Distinguishing between directly and indirectly light-activated neurons is often non-trivial, especially within circuits with recurrent connections [38, 328, 465], and direct validation is challenging because extracellular recording methods do not provide access to anatomic information of the recorded neurons. These limitations highlight the need for 'ground truth' data obtained from individual, anatomically identified LC neurons.

To bridge this gap, we employed the juxtacellular technique for recording and labelling single neurons *in vivo* in the mouse LC complex. Specifically, we aimed to characterize the *in vivo* activity of identified noradrenergic neurons and asked whether electrophysiological properties alone can be sufficient for distinguishing them from the heterogeneous surround of non-noradrenergic cells. Based on our dataset of identified neurons ( $n = 50$ ), we describe *in vivo* electrophysiological correlates of noradrenergic neurons within the mouse LC complex.)

## Results

To characterize the electrophysiological and morphological properties of LC neurons, we performed juxtacellular recording and labelling of single neurons in anaesthetized mice (**Figure 3.13a**). The cytochemical identity of the labelled neurons was assessed via immunoreactivity for tyrosine hydroxylase (TH)—the rate-limiting enzyme in catecholamine biosynthesis, and a classical marker for LC noradrenergic neurons [27, 69, 177, 242, 306, 346]. In line with previous work,

### 3. RESULTS

---

TH-positive (putatively noradrenergic) neurons were most abundant at posterior-dorsal levels of the LC (classically referred to as the LC ‘nuclear core’, [Figure D.1](#)); however, outside of the LC ‘core’, TH-positive neurons were more intermingled with putatively non-noradrenergic (TH-negative) cells—the latter being identified by the lack of immunoreactivity for TH, and the presence of immunoreactivity for the pan-neuronal marker NeuN, and DAPI (Figures 1b and S1). These observations are in line with previous work on the rodent LC, which indicated that at the antero-ventral aspect of the LC complex, scattered noradrenergic neurons intermingle within the lateral dorsal tegmental nucleus (LDTg) and/or central grey [146, 390]. In this study, we did not restrict our analyses to the LC ‘nuclear core’ but instead focused on this extended LC complex ([Figure D.1](#); for review, see [242]).

[Figure 3.13c-h](#) shows a representative recording from an identified LC neuron that was positive for TH immunoreactivity (TH-positive; [Figure 3.13c](#)). This neuron displayed a multipolar morphology, and its dendritic tree sampled a broad territory outside the area containing TH-positive neuronal somatas ([Figure 3.13d](#)). Furthermore, this TH-positive neuron fired at a relatively low rate (0.51 Hz) and displayed a broad spike waveform ([Figure 3.13f](#)) and a low tendency to fire spikes at short interspike intervals ([Figure 3.13g](#))—in line with previous observations [18, 395, 405, 406, 438]. Furthermore, this TH-positive neuron exhibited a multiphasic excitatory response to foot shock stimulation, consisting of a short-latency component, followed by additional activity peaks that outlasted the stimulus itself [73, 163], followed by post-activation inhibition [5, 70, 114] ([Figure 3.13h](#)). Notably, this neuron reliably responded to the stimulation, with 109 out of 109 stimulation trials displaying an evoked spike within a 100 ms window following stimulus onset (Response Reliability = 1, see methods section [D.2](#)).

[Figure 3.13i-n](#) shows a representative example from an identified TH-negative

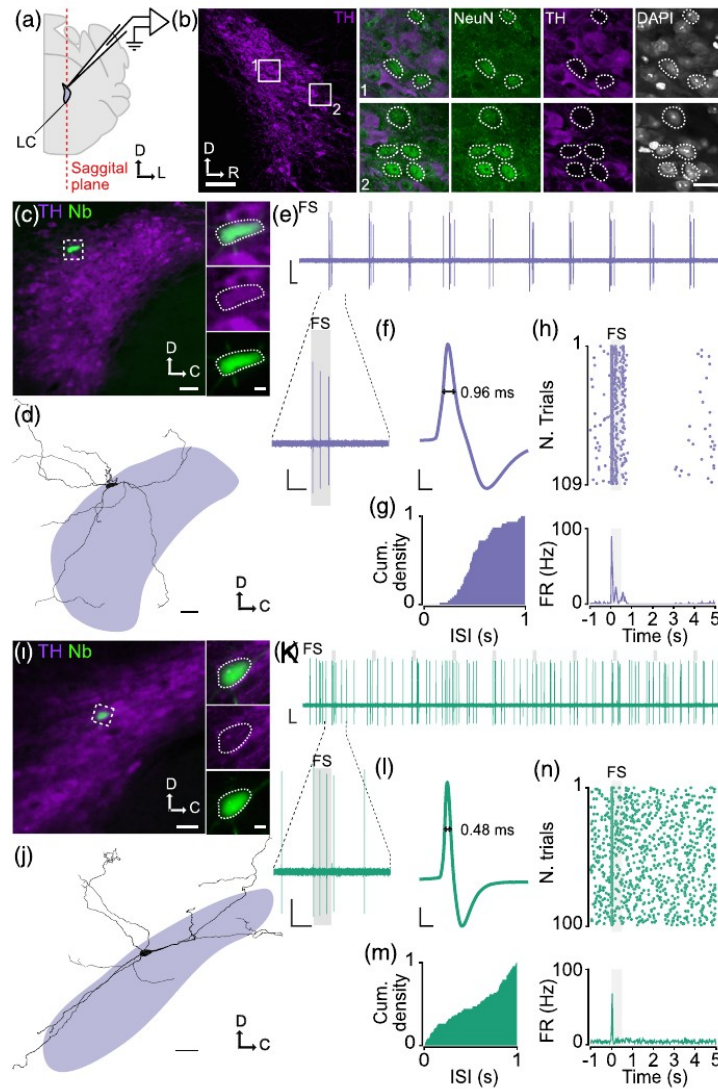


Figure 3.13: **In vivo recordings from anatomically identified TH-positive and TH-negative neurons within the mouse LC complex.** (a) Schematic of the experimental preparation, showing juxtacellular recording in the mouse LC. The parasagittal sectioning plane for post hoc histological analysis is indicated with a red dotted line. D, dorsal; L, lateral; LC, locus coeruleus complex. (b) Fluorescence images of a parasagittal tissue section through the LC (as in a). Left, staining for tyrosine hydroxylase (TH; magenta). Scale bar, 100  $\mu\text{m}$ . Right, high-magnification insets showing staining for NeuN (green), TH (magenta) and DAPI (white). TH-negative cells, positive for NeuN and DAPI, are indicated with dotted lines. Scale bar, 20  $\mu\text{m}$ . D, dorsal; R, rostral. (c) Left, fluorescence image of a parasagittal section through the mouse LC, stained for TH (magenta), showing a representative identified TH-positive neuron (green; Neurobiotin, Nb) recorded in vivo. Scale bar, 50  $\mu\text{m}$ . Right, high-magnification detail corresponding to the area indicated with a dotted line in the left image, showing the neuron's soma (green; Nb) colocalized with TH (magenta). Scale bar, 10  $\mu\text{m}$ . (d) Morphological reconstruction of the neuron shown in a, superimposed on an outline of the TH fluorescence signal (magenta). Scale bar, 50  $\mu\text{m}$ . (e) Top, representative high-pass filtered juxtacellular voltage trace from the neuron shown in c, d, recorded during foot shock stimulation (FS). Note the reliable spiking in response to the stimuli (indicated in grey). Bottom, high-magnification view on a single stimulation trial (dotted box). (f) Average spike waveform for the identified neuron shown in c, d. The spike half-width is indicated. (g) Cumulative distribution of interspike intervals (ISI) during spontaneous activity for the representative neuron shown in c, d. (h) Raster plot (top) and peristimulus time histogram (bottom) during foot shock stimulation (grey area) for the neuron shown in c, d. (i-n) Same as c-f, but for a representative TH-negative neuron. Scale bars in i, 50  $\mu\text{m}$  (left), 10  $\mu\text{m}$  (right). Scale bar in j, 50  $\mu\text{m}$ .

### 3. RESULTS

---

neuron. Like the TH-positive cell (**Figure 3.13d**), this neuron displayed a multipolar morphology (**Figure 3.13j**) and fired at a relatively low rate (0.73 Hz). However, compared with the TH-positive cell (**Figure 3.13f**), this neuron displayed a narrower spike waveform (**Figure 3.13l**) and distinct temporal firing patterns, characterized by a higher tendency to fire at short interspike intervals (**Figure 3.13m**). Notably, this TH-negative neuron also reliably responded to the foot shock stimulus (**Figure 3.13n**) with high trial-to-trial reliability (Response Reliability = 0.87, i.e. 87 out of 100 response trials with an evoked spike within a 100 ms window following stimulus onset, **Figure 3.13n**; see methods section **D.2**).

Altogether, we recorded 50 identified neurons in the extended LC complex of anaesthetized mice (see **Figure D.2**). Of these neurons, 60% (30 out of 50) were found to be positive and 40% (20 out of 50) negative for TH immunoreactivity, consistent with the relative abundance of these cell types from anatomical analysis (**Figure D.1C**). A subset of neurons with high-quality filling (TH positive,  $n = 13$ ; TH-negative,  $n = 11$ ; see methods section **D.2**) was reconstructed for morphological analysis. On average, TH-negative and TH-positive neurons displayed largely overlapping morphometric features (**Figure D.3**), indicating that neurochemical identity (i.e. TH expression) cannot be directly inferred from morphological criteria alone. Nevertheless, we acknowledge the possibility that due to the relatively small data-set of reconstructed neurons ( $n = 24$ ), possible (subtle) morphological differences between the two cell classes might have gone undetected.

Previous studies differentiated heterogeneous LC neurons by their average spontaneous firing rate and spike width, which motivated us to compare these characteristics between TH-positive and TH-negative neurons (e.g. [404, 405]). On average, compared with TH-positive neurons, TH-negative cells displayed lower spontaneous mean firing rates (TH-pos,  $1.21 \pm 0.78$  Hz; TH-neg,  $0.77 \pm 0.62$  Hz;  $p =$

0.036; **Figure 3.14a**) and narrower spike waveforms (spike half-width durations;  $0.83 \pm 0.18$  Hz; TH-neg,  $0.69 \pm 0.14$  Hz;  $p = 0.008$ ; **Figure 3.14b,c**). Moreover, TH-positive neurons exhibited a higher tendency to fire spikes at short ISI (**Figure 3.14d,e**). Notably, in most neurons, we observed the presence of a ‘shoulder’ on the descending phase of the juxtacellular spike waveform (**Figure D.4**). As shown in other cell types [102], the presence of a shoulder resulted in relatively long peak-to-trough spike durations in our dataset (see **Figure D.4**). These observations are consistent with prior extracellular recordings of LC units, which also showed ‘broad’ spike waveforms featuring a ‘shoulder’ in the descending phase of the spike, as well as a ‘notch’ in the ascending phase (**Figure D.4C**; e.g. [21, 431, 438]).

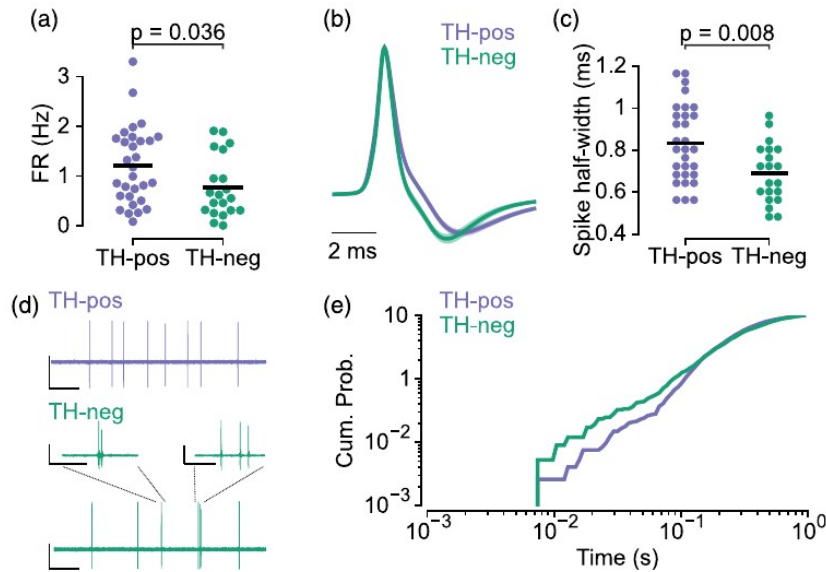


**Figure 3.14: Electrophysiological properties of TH-positive and TH-negative neurons.** (a) Average spontaneous firing rates of TH-positive ( $n = 30$ ; magenta) and TH-negative ( $n = 20$ ; green) neurons. Lines indicate mean values. P-value is indicated (Mann–Whitney–Wilcoxon test, two-sided). (b) Superimposed average spike waveforms for TH-positive and TH-negative neurons ( $n$  and conventions as in a). Shaded regions indicate the standard error of means. (c) Spike half-width durations for TH-positive and TH-negative neurons ( $n$  and conventions as in a). (d) Representative juxtacellular spike traces from TH-positive (magenta) and TH-negative (green) neurons recorded during spontaneous activity. Insets show high magnification of a spike-double and -triplet from the TH-negative neuron. Scale bars, 1 mV, 1 s and 100 ms (insets). (e) Mean cumulative distributions of interspike intervals for TH-positive and TH-negative neurons ( $n$  and conventions as in a).

To further investigate the differences in electrophysiological properties between

### 3. RESULTS

---

TH-positive and TH-negative neurons, we extracted temporal features of spike waveforms, interspike interval distributions and spike autocorrelograms using sPCA (**Figure 3.15A-C**; see methods section **D.2**). To validate differences in in vivo electrophysiological properties between TH-positive and TH-negative neurons, we undertook two independent approaches. First, we performed PCA on all LC neurons using spike-waveform and spike-train metrics (see **Figure 3.15D-F** and Section 2). Indeed, TH-positive and TH-negative cells were partially separated along the two first principal dimensions (**Figure 3.15E**). Second, a logistic-regression classifier trained on the same electrophysiological features was able to separate TH-positive and TH-negative neurons with reasonable accuracy (cross-validated accuracy: 76%), with the largest weights being assigned to spike waveform features and spontaneous mean firing rate (**Figure 3.15G-H**). Altogether, these data indicate that TH-positive neurons display distinct spontaneous electrophysiological features compared with putative non-adrenergic cells.

In a subset of recordings ( $n = 23$  out of 50; TH-positive,  $n = 14$ ; TH-negative,  $n = 9$ ; **Figure D.1**), we delivered foot shock stimuli consisting of individual pulses at similar intensity (1 mA) but relatively longer duration (0.5 s) compared with previous studies (e.g. [70, 73, 284, 405]). Previous work has shown that short-duration stimuli (0.5-5 ms) typically evoke a biphasic response in LC neurons, consisting of a phasic excitatory response followed by transient post-activation inhibition (e.g. [19, 73, 284, 405]); the latter suppression is classically attributed to noradrenergic-mediated mechanisms, and its duration is known to be influenced by the stimulus characteristics [5, 70, 114]. Under our stimulus regime, both TH-positive and TH-negative neurons displayed excitatory foot shock responses, characterized by a phasic component, usually followed by additional activity peaks (Figures 3a and S6; see also [73, 163]). On average, the activity of TH-positive neurons showed stronger foot shock modulation than that of TH-negative cells, as indicated by the

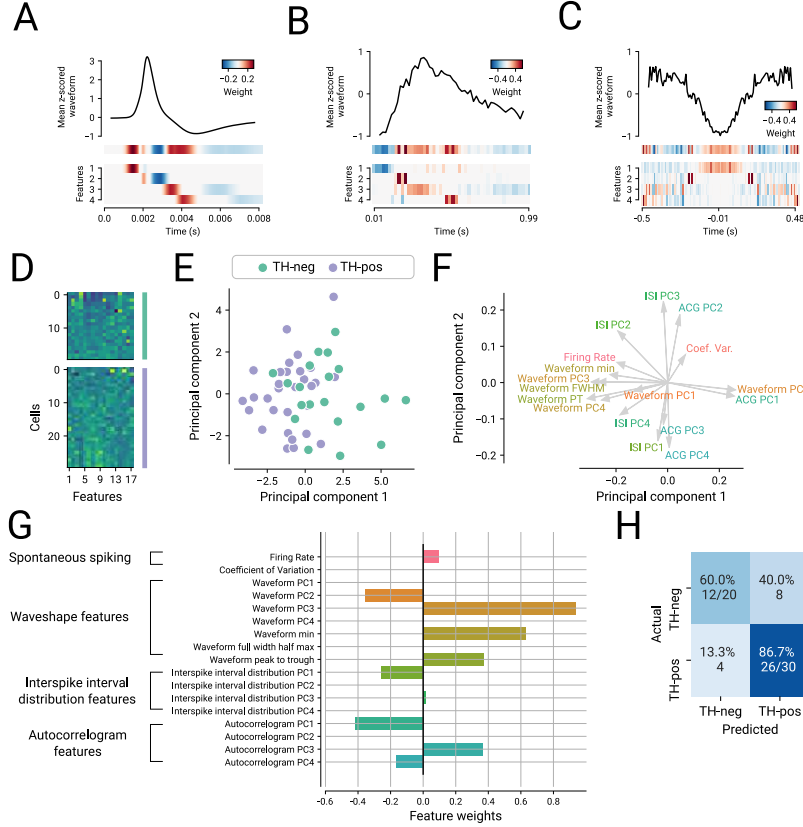


**Figure 3.15: Separating TH-positive and TH-negative cells using sparse PCA and a logistic regression classifier.** (A-C) Features of temporal electrophysiological characteristics of LC neurons were extracted using sparse PCA (sPCA). (A) Top: mean z-scored spike waveform across all neurons (black, n=50). Bottom: temporal features extracted by sPCA. (B) Same as in A, but for interspike interval distributions. (C) Same as in A, but for autocorrelograms. (D) Feature matrix of TH-positive neurons (n=30; magenta) and TH-negative neurons (n=20; green), and electrophysiological spike-waveform and spike-train metrics (nfeatures=17; see G, left). The feature matrix is sorted by the first principal component (see E). (E) To confirm the differences in in vivo electrophysiological properties between TH-positive and TH-negative neurons, we used a two-dimensional PCA embedding of TH-positive (magenta) and TH-negative (green) neurons. Indeed, TH-positive and TH-negative neurons were partially separated along the two first principal dimensions. (F) PCA biplot showing the relative contribution of each feature to the two principal components (i.e. the length of the arrows) and the relationship between features (i.e. arrows with a common direction show correlations between features). Waveform PT, Peak-to-trough duration; ACG, Autocorrelogram; ISI, Interspike interval; PC, principal component. (G) Feature weights of the logistic regression classifier. The classifier was trained to separate TH-positive and TH-negative cells based on their electrophysiological properties. We used a combination of features (shown in D) that described the following characteristics: spontaneous firing (num. features = 2), the spike waveform shape (num. features = 7), the interspike interval distribution (num. features = 4), and the spike autocorrelogram (num. features = 4). Same color codes as in F. (H) Confusion matrix for the logistic regression classifier.

### 3. RESULTS

---

higher foot shock Modulation Indices (see methods section D.2; TH-positive,  $0.9 \pm 0.07$ ; TH-negative,  $0.71 \pm 0.2$ ;  $p = 0.009$ ; **Figure 3.16b**) and Response Ratios (TH-positive,  $89.38 \pm 61.27$ ; TH-negative,  $44.27 \pm 44.87$ ;  $p = 0.021$ ). On the other hand, response reliability (see methods section D.2; TH-positive,  $0.74 \pm 0.25$ ; TH-negative,  $0.8 \pm 0.2$ ;  $p = 0.899$ ; **Figure 3.16c**) and response latencies (see methods section D.2; TH-positive,  $51.83 \pm 13.6$  ms; TH-negative,  $47.21 \pm 16.79$  ms;  $p = 0.469$ ) were similar between the two cell classes. Notably, in a subset of cells, our foot shock stimuli induced a prolonged post-activation suppression of firing that outlasted the typical duration observed with shorter-duration stimuli (**Figure 3.16d-f**; see also [5, 70]). On average, the spiking activity of TH-positive neurons was more suppressed than that of TH-negative cells (average firing rate in interstimulus intervals: TH-positive,  $0.14 \pm 0.13$  Hz; TH-negative,  $0.69 \pm 0.51$  Hz;  $p = 0.013$ ; post-activation suppression index: TH-positive,  $-0.63 \pm 0.34$ ; TH-negative,  $-0.18 \pm 0.5$ ;  $p = 0.015$ ; see methods section D.2; **Figure 3.16e-f**), indicating that noradrenergic-mediated auto- and collateral inhibition—features classically attributed to LC cells [5, 70, 114] - could potentially contribute to the observed post-activation suppression of activity. We note, however, that our conclusions rest on a relatively low number of identified neurons; hence, future work will be required to extend these observations and explore possible differences in the temporal dynamics of foot shock responses between the two cell classes.

One limitation of the current dataset is that the identified neurons were classified according to the expression of a single marker (i.e. TH immunoreactivity). TH expression in the LC has been shown to be activity-dependent, at least under certain experimental conditions (e.g. [294, 462]), thereby raising the possibility of false-negative TH classification in our dataset. Moreover, the biological variability of TH expression levels among LC neurons might also reduce confidence in cell identification, especially in cells with relatively low TH expression. To ad-

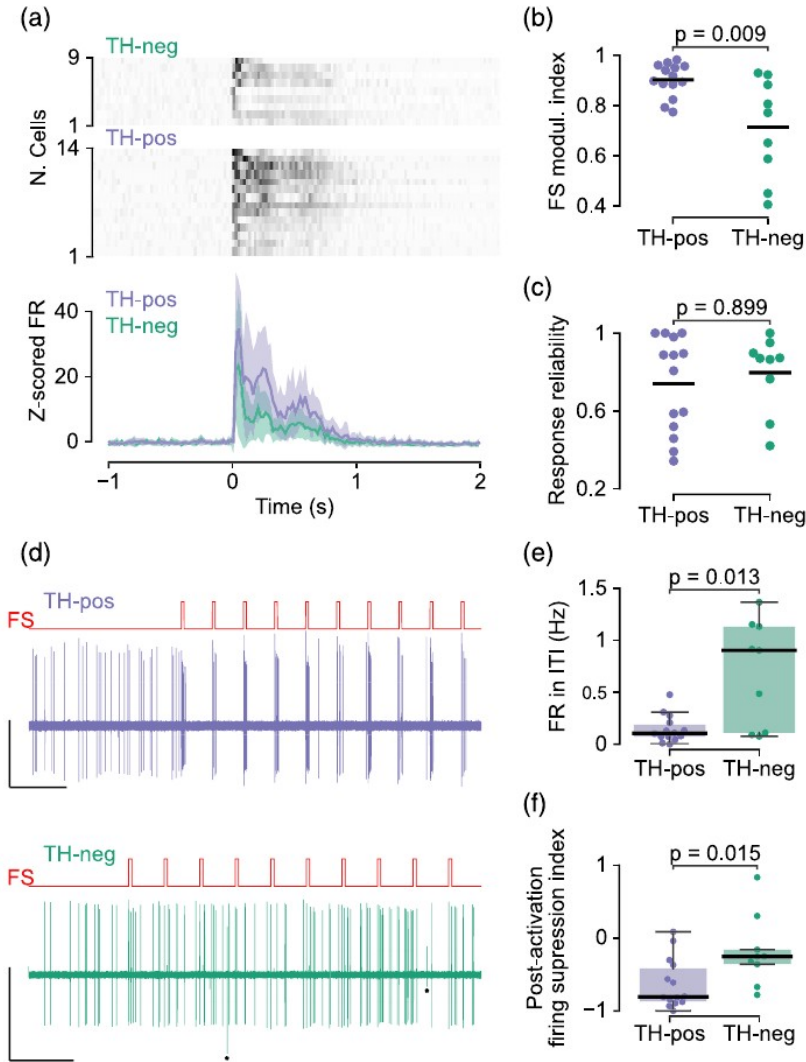


Figure 3.16: **Foot shock responses of TH-positive and TH-negative neurons.** (a) Top, z-scored peristimulus-time activity of TH-positive ( $n = 14$ ) and TH-negative ( $n = 9$ ) neurons during foot shock stimulation (FS; grey). Bottom, average z-scored responses from the peristimulus activity shown above. Shadows indicate the standard deviation. (b) Foot shock Modulation Indices for TH-positive ( $n = 14$ , magenta) and TH-negative ( $n = 9$ , green) neurons. Mean values are denoted by lines. P-value is indicated (Mann–Whitney–Wilcoxon test, two-sided). (c) Foot shock response reliability for TH-positive and TH-negative neurons. Same  $n$  and conventions as in b. (d) Representative high-pass filtered juxtacellular voltage traces from a TH-positive neuron (top) and a TH-negative neuron (bottom) during FS (red). Note the long-lasting post-activation suppression during inter-trial intervals (ITI) for the TH-positive neuron. Asterisks indicate artefacts. Scale bars, 2 mV, 10 s. (e) Average firing rates during ITIs for TH-positive and TH-negative neurons. Same  $n$  and conventions as in b. (f) Post-activation suppression index (see methods section D.2), which quantifies the degree of firing rate suppression in ITIs compared with baseline (spontaneous firing), for TH-positive and TH-negative neurons. Same  $n$  and conventions as in b.

### 3. RESULTS

---

dress these concerns, we first stained the LC for additional noradrenergic markers, DBH and NET [120, 146, 235, 365, 390, 403, 418], and quantified the degree of colocalization with TH (**Figure D.6**). The large majority of DBH-positive and NET-positive cells were also found to be immunoreactive for TH (TH and NET double-positive neurons:  $98.15 \pm 2.12\%$ ;  $n = 1489$  neurons,  $n = 5$  nuclei; TH and DBH double-positive neurons:  $96.96 \pm 1.83\%$ ;  $n = 1614$  neurons,  $n = 5$  nuclei, **Figure D.6**), thus pointing to a relatively low probability of false TH-negative neurons in our dataset. Second, restricting the electrophysiological analysis to the cells with ‘high confidence’ of immunohistochemical classification (based on the overall quality and intensity of the fluorescence signal) led to qualitatively similar results regarding spontaneous electrophysiological properties of TH-positive and TH-negative neurons (see methods section D.2; not shown). Altogether, this evidence indicates that our results are not significantly biased by potential misclassification of neuronal identities based on TH expression.

### Discussion

Brainstem noradrenergic neurons, through their wide-spread projections to the rest of the central nervous system, are in a privileged position for controlling a wide range of physiological brain functions, including arousal, attention and memory [20, 117, 315, 338, 406]. Despite their crucial importance, the small size and deep location of this neuronal population and the lack of clear electrophysiological correlates of anatomical boundaries have made the *in vivo* characterization of noradrenergic single-unit activity particularly challenging.

In the present study, we aimed at bridging this gap by taking advantage of the *in vivo* juxtacellular recording-labelling method. Specifically, we applied this technique to a preparation that has been extensively used to study the *in vivo* activity of the LC, namely, foot shock stimulation in anaesthetized animals (e.g.

[73, 74, 284, 334, 405]. This approach offers two main advantages compared with conventional extracellular techniques. Firstly, the high signal-to-noise ratio of juxtacellular spike signals enabled the precise identification of spikes, even during highly synchronous LC population activity, thus providing accurate ('ground truth') spiking patterns from individual LC cells. Secondly, by labelling recorded cells and subsequent immunohistochemical and morphological analysis, we could link in vivo recorded activity patterns to individual neurons.

The main goal of the study was two-fold: first, to characterize the in vivo activity of identified noradrenergic neurons within the mouse LC complex; second, to explore whether these properties can be used for separating noradrenergic cells from the heterogeneous surround of non-noradrenergic neurons. We found that noradrenergic neurons (identified by the expression of TH, TH-positive) displayed significantly different spontaneous electrophysiological features (i.e. mean firing rates (**Figure 3.14a**), spike waveforms (**Figure 3.14b,c**), temporal spiking properties (**Figure 3.14d,e**) and average foot shock responses (**Figure 3.16b**) compared with putative non-noradrenergic (TH-negative) cells. Spontaneous in vivo electrophysiological features alone were sufficient for classifying TH-positive and TH-negative cells with 76% accuracy (see **Figure 3.15G-H**). These observations are consistent with recent studies, which also describe neuronal subpopulations with relatively narrow spike waveforms in the mouse, rat and primate LC [386, 404, 405]. Morphologically, TH-positive and TH-negative neurons displayed largely overlapping features (**Figure D.3**), which prevented cell type classification based solely on morphological criteria. Notably, we observed a high degree of morphological variability even within the same class (**Figure D.3**), pointing to further heterogeneity waiting to be resolved.

One limitation of the current study relates to the definition of noradrenergic and non-noradrenergic cell types based on TH immunohistochemistry. Although this

### 3. RESULTS

---

approach has been validated and extensively employed in previous studies of the rodent and human LC complex [27, 69, 177, 242, 306, 346], it is known that TH expression can be modulated by neuronal activity [294, 462]. Therefore, since TH-negative neurons in our dataset were defined by the ‘absence’ of marker expression, false negatives arising from technical sensitivity and/or biological variability cannot be formally ruled out. However, colocalization analysis with additional noradrenergic markers (DBH and NET) revealed a high degree of consistency with TH expression (**Figure D.6**), indicating that under our experimental conditions, the occurrence of false TH-negative cells is likely to be very low. Moreover, restricting the analysis to cells with ‘high confidence’ of immunohistochemical classification led to qualitatively similar results (see methods section D.2), indicating that our conclusions are unlikely to be significantly biased by inadvertent cell misclassification.

The relative distribution of TH-positive and TH-negative cells within the mouse LC complex was consistent with previous results. Specifically, TH-positive neurons were most abundant within the densely packed ‘nuclear core’ of the LC, where TH-negative cells represented a small minority of all neurons (**Figure D.1B,C**; [242]). On the other hand, noradrenergic cells were more scattered—and hence more intermingled with TH-negative neurons—outside of the ‘nuclear core’ region (e.g. at more antero-ventral levels, where intermingling of noradrenergic neurons with LDTg and/or central grey has been described [132, 146, 390]. Since different non-noradrenergic cell types with various neurotransmitter phenotypes (e.g. GABAergic, cholinergic and glutamatergic) have been described within (and especially around) this most antero-ventral aspect of the rodent LC complex [22, 50, 85, 132, 139, 182, 207, 237], our TH-negative group likely includes heterogeneous subpopulations of neurons. For example, TH-negative neurons (**Figure D.2**) could possibly also include LDTg neurons, which have been

shown to exhibit broad action potentials and respond to sensory stimuli [203]. In the present study, these possibly diverse TH-negative neurons were grouped in order to test whether TH-positive neurons display unique electrophysiological features compared with the heterogeneous surround of non-noradrenergic cell types. We also note that noradrenergic neurons have also been shown to be heterogeneous [71, 263, 404, 405], but our limited sample size prevented further within-class analysis. Future work will be required to resolve additional structure-function relationships within these heterogeneous TH-negative (and TH-positive) cell classes, which was beyond the scope of the present study.

In this study, in vivo recorded and identified TH-positive neurons were sampled from a region extending beyond the densely packed LC ‘nuclear core’ (Figures S1A and S2). Whether noradrenergic subpopulations within this extended LC complex are functionally distinct remains to be demonstrated. Our relatively small dataset, together with the difficulty of unequivocally assigning neurons to specific subregions of the LC complex, particularly in parasagittal sections, prevented the analysis of possible relationships between the anatomical location of the recorded cells and in vivo activity patterns. We envision that the registration of neuronal morphologies onto a standardized anatomical reference frame of the LC complex—following the approach being established for the neocortex (e.g. [210, 290])—will enable the integration of data collected from different studies and laboratories and will be instrumental for resolving possible structure-function relationships at the single cell level.

Another limitation of this study is the use of xylazine, an alpha-2 adrenergic receptor agonist, in the anaesthetic mixture (see methods section D.2; [4]). However, in our preparation, several electrophysiological features of LC activity could be recapitulated. For example, spontaneous firing rates (Figure 3.14a), spike waveform (Figure 3.14b,c and S4), spike train features (Figure 3.14d,e) and foot shock

### 3. RESULTS

---

responses (Figures 3 and S6) were consistent with estimates from prior work based on different anaesthetics or species (e.g. [73, 395, 405]). This indicates that several electrophysiological features classically associated with LC neurons were preserved in our preparation. We acknowledge, however, that despite similarities, a direct comparison between studies using different anaesthetics and/or animal species (e.g. rats vs. mice) is not entirely warranted because species- and/or anaesthetic-specific effects cannot be excluded a priori. For example, while spontaneous firing rates in our study were largely consistent with estimates from previous in vivo work [18, 73, 405, 406], Sugiyama [387] reported higher firing rates under a similar recording configuration (in vivo cell- attached recordings) but in urethane anaesthetized rats. We also acknowledge that the use of xylazine, which acts on  $\alpha_2$ -adrenoceptors, might have suppressed the firing rates of noradrenergic cells in our study and, hence, narrowed the difference in excitability between the TH-positive and TH-negative groups. Future in vivo recordings in drug-free animals will be required to extend our observations to the natural system during awake behaviour.

The advent of optotagging techniques in the study of LC circuits has provided an unprecedented level of spatial and temporal resolution (for review, see [397, 406]). However, these systems-level approaches are not optimally suited for resolving structure-function relationships, as they are typically restricted to targeting ‘broad’ cell types (i.e. defined by single marker expression) and do not provide access to morphological information. Additionally, known limitations of spike-sorting algorithms, particularly during epochs of highly synchronous activity [128, 131, 223, 329] such as the phasic response to foot shocks, have hindered the precise characterization of LC activity patterns. Here, we provide, to the best of our knowledge, the most extensive dataset of in vivo recorded and morphologically identified brainstem noradrenergic neurons. Our juxtacellular record-

ings from single, anatomically identified neurons provide ‘ground-truth’ firing patterns, which can be useful for benchmarking and/or validating current spike-sorting-based approaches for the classification of noradrenergic unit activity.

Altogether, our dataset of identified neurons reveals in vivo electrophysiological correlates of noradrenergic neurons in the mouse LC complex and contributes to the ongoing research efforts aiming at the multimodal characterization of brain-stem noradrenergic neurons.

## 4 | DISCUSSION

In this thesis, I presented how statistical methods and computational modelling can be used to gain insights into fundamental brain mechanisms. This included the analysis of functional, structural, and molecular information of neurons in different brain areas and at different scales (single neurons *versus* neuronal populations). Particularly, I focused on the question of how sensory and behavioral inputs are integrated and represented in the brain and how this recurrently influences the animal's behavior and neural plasticity depending on internal states (e.g. arousal) and the external context (e.g. the stimulus).

### 4.1 Summary

Study I (3.1) investigated feedforward and feedback processing in the cortico-thalamic pathway. By extending a data-efficient version of generalized linear model, we predicted extracellular responses of neurons in the dLGN in awake mice. We found that cortico-thalamic (CT) feedback and behavioral states, indexed by locomotion and pupil size, significantly influenced dLGN activity. Particularly, we identified one group of neurons that are significantly modulated by non-retinal inputs, and that was poorly responsive to visual stimuli otherwise. Further, we found that the responses of CT feedback-sensitive dLGN neurons depended on the type of visual stimulus and contained more visual information when CT feedback was suppressed. Together, our results show that the integration of visual information depended on the behavioral context and stimulus conditions.

In Study II (3.2) and Study III (3.3), I presented an investigation neural mechanisms underlying functions in the lateral habenula (LHb) using data from in vivo imaging and single-unit recordings in mice. Study II characterized the electro-

physiological diversity of LHb neurons and their unique responses to noxious stimuli and behavioral states using an unsupervised learning technique. Through this approach, we identified distinct neural firing patterns and their correlations with stimulus sensitivity and spontaneous behaviors in both anesthetized and awake conditions. Together, this provided insights into the cellular heterogeneity of the LHb and its role in integrating aversive information to modulate behavior. Study III investigated the neural mechanisms underlying threat association in the LHb. We found that neurons in the mouse LHb reveal both, excitatory and inhibitory responses to aversive stimuli. Further, we found that populations in the LHb dynamically encoded associative learning between conditioned and unconditioned aversive stimuli. We also identified specific synaptic changes and neurotransmitter dynamics, indicating a complex interplay of presynaptic and postsynaptic mechanisms during cue-punishment learning.

Finally, in Study IV (3.4) I analysed juxtacellular recordings from identified neurons in the mouse locus coeruleus (LC), a hub for noradrenergic transmission. We found that neurons molecularly identified as either noradrenergic or non-noradrenergic within the LC have distinct electrophysiological properties. Using an unsupervised logistic regression classifier, we were able to distinguish between the two molecular cell classes with 76% accuracy, only trained on spontaneous electrophysiological features. Taken together, our results show the complexity of neurotransmitter-specific neural dynamics and its implications for stress and arousal regulation.

## **4.2 The challenge of balancing detail and utility**

When discussing statistical frameworks in the context of neuroscience, it is important to think about the right level of abstraction. This applies to model complexity

## 4. DISCUSSION

---

(i.e. what details are included or ignored in the model), as well as finding the right scale (i.e. the inherent uncertainty in determining the exact number of clusters without ground truth). I will discuss these issues in the following two paragraphs and give some insights into potential solutions.

### **Model complexity**

A statistical model is an abstract representation of the real world system. In Study I (3.1), I used an efficient version of a generalized linear model (GLM): each thalamic neuron was represented by one simple GLM model. This is a highly simplified abstraction of the real world system with all its molecular detail. On the other extreme, especially in vision research, deep neural networks (DNNs) are often used to explain the activity of large populations of neurons. Here, each layer in the network can be associated to a processing stage in the visual pathway of the brain and one model representing hundreds of neurons at the same time [191, 192, 421, 446]. Even though these DNN models are undefeated in accuracy when predicting neural activity, it would not be feasible with the hardware available currently and would be computationally too expensive to capture every molecular detail of the brain<sup>1</sup>. Consequently, finding the right level of abstraction for the scientific question one is investigating is an essential part of computational modelling. George Box captured this by saying: “all models are wrong, but some are useful” [49]. Somewhat less known, however, is how the famous sentence of George Box continues: “However, the approximate nature of the model must always be borne in mind.” In other words, the right level of abstraction needs to be chosen by the person who creates the model that is inherently an abstraction of the

---

<sup>1</sup>Significant progress has been made over the past years to develop mechanistic models of single neurons and neural populations that are computationally efficient while still capturing substantial molecular details [99, 185]. However, as discussed in 1 Introduction, this thesis focuses on phenomenological models. Therefore, mechanistic models are not further discussed here.

real world system. Consequently, the right model is the one with the least detail that still integrates the functionality we are interested in. One strategy to find the right level of abstraction is to start with some plausible level of abstraction and then keep removing details until the behavior of interest is not captured anymore [67]. Cao and Yamins [67] argue that once we reach this point, we have found a model that shares the same mechanism as the biological system.

One well-known framework to understand computational systems such as the brain are the three levels of abstraction by David Marr [250]:

1. *Computational level*: What is the goal of the computation and what is the problem it solves?
2. *Algorithmic level*: What are the steps carried out by the system to produce the output?
3. *Implementational level*: How are the computations physically implemented by the hardware?

Applied to the corticothalamic circuit discussed in Study I (3.1), one could identify the three levels as follows: The *computational level* would be the ability of the system to process visual information and creating a coherent representation of the visual world. This allows the organism to interpret and interact with its environment effectively (e.g. object recognition, motion detection, and spatial awareness). The *algorithmic level* would be the processing of visual information steps at the retina (capturing light and converting it into a neural signal), which is then send to the visual thalamus. This is followed by further processing in the primary visual cortex, where the information is integrated, certain features are enhanced and identified to create a detailed map of the visual scene. The *implementational level* would then be how the computations are physically realized by neurons and circuits in the retina, the visual thalamus, and primary visual cor-

## 4. DISCUSSION

---

tex. Particularly, the implementational level would study how these neural mechanisms involve synaptic transmissions and action potentials that travel through the optic nerve to the thalamus and the primary visual cortex, where cortical layers contain specialized neurons to process specific visual attributes. Even though the model in Study I (3.1) is a simple abstract representation of the neural circuit and modulatory inputs, it is well interpretable. The shapes and the amplitude of the stimulus kernel, the behavioral kernels and the corticothalamic-feedback kernel are representative for the strengths and the temporal dynamics with which the inputs modulate the neural activity. Finding the simplest model that explains the data well among competing models is the core principle of Occam’s razor: if a model becomes too complex, it overfits to the data and captures noise instead of signal [153].

### **Finding the sweet spot: clustering of neuronal cell types**

Another aspect in which finding the right scale matters is in clustering, which was relevant in Study II (3.2) and Study III (3.4). Visualizing observations using t-SNE often shows that neurons within the same cluster can still show variability due to biological and technical reasons. How can we find the ideal split into groups or neuronal cell types? From the statistical point of view, there are measures such as the Bayesian information criterion (BIC), which optimise the model complexity (i.e. number of clusters) and the ability of the model to best explain the data [348].

Focusing on discontinuous features helps to define clear groups, while continuous feature variations create a conceptual problem for deciding how coarse or detailed the distinction into types should be. To acknowledge the feature gradients across cell types rather than the existence of distinct, clearly separable clusters, Zeng et al. (2017) [455] suggests looking at cell type classification from a hierarchical perspective, where all degrees of granularity coexist, just at different hierarchical

ranks. This allows for a fine grained and a coarse grouping existing in parallel: the smallest functional entity should be called a type and the grouping of multiple types is called a class. In between type and class are subclasses. Therefore, one neuron can belong to multiple classifications at the same time depending on the level. This makes it possible to split groups or interpolate between groups as new information becomes available.

In the future, it will be important to explore alternatives to classical approaches such as probabilistic clustering (to return uncertainty measures for one neuron being part of a particular group) or bi-clustering (to integrate multiple modalities into the pipeline). Further, the technical advances will scale up the number of simultaneously recorded neurons and the size of datasets. This trend towards 'big data' will require more automatic feature extraction methods and clustering algorithms not only for hand-tuned datasets. The goal should be to extend clustering to global algorithms involving multiple domains combining functional, morphological and molecular information.

### **4.3 Our brains, our future?**

The potential synergies between neuroscience and statistics are broad. The recent progress in machine learning (as a sub-field of statistics) and artificial intelligence (AI) in combination with large-scale data allows the usage of statistical tools at an unprecedented scale: on the one hand, advances statistics and modelling drive the discovery of brain functions and underlying mechanisms [1, 134]; on the other hand, insights from neuroscience also have the potential to inspire the development of new machine learning tools or AI systems with improved robustness and flexibility [151]. Further, AI models such as DNNs have too many parameters to be interpretable in a straightforward way and are too complex to be

## 4. DISCUSSION

understood directly. To tackle this problem, tools used in neuroscience can be applied to understand how information is represented in AI models, e.g. by creating and analysing low dimensional representations of weights or layer activity. The study of biological and artificial intelligence is referred to “NeuroAI” (Figure 4.1, [267]). This highly interdisciplinary and still emerging research field has the goal to find universal principles for intelligence in both, biological and artificial agents. Within "NeuroAI", various research streams and approaches exist, depending on what motivates the research. Both perspectives are present: using insights from neuroscience to improve AI systems (Figure 4.1 top left) and applying AI models to explain neural data (Figure 4.1 bottom right)<sup>2</sup>.

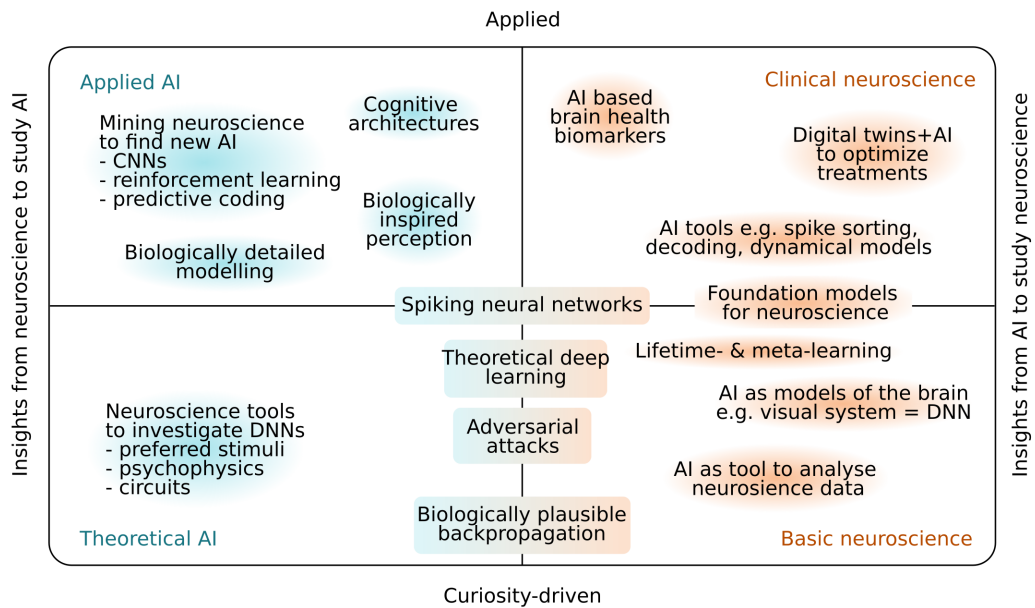


Figure 4.1: **Topology of "NeuroAI"**. Schematic overview of the different subfields in "NeuroAI". The horizontal axis describes the two application areas: AI (*left*) and neuroscience (*right*). On the vertical axis, the topics are split into applications (*top*) and curiosity-driven (basic) research (*bottom*). This figure was adjusted from Mineault (2023) [267].

One key goal of "NeuroAI" is to create simulations that reproduce the observed neural activity in order to study and understand the neural system better [108,

<sup>2</sup>The research contained in this thesis would be located at the *bottom right* in Figure 4.1: using statistical frameworks and AI tools to investigate neuroscience data.

453]. The idea is that if the statistical model generates the same output as the biological systems, it will be easier to understand the underlying mechanisms because the computational model is easier to analyse and dissect. However, this approach assumes a single solution, whereas multiple different mechanisms can produce the same output [246]. This is known as the “many-to-one mapping problem”. With the rapid progress in acquiring large-scale data sets, one recent trend in computational neuroscience has been to focus on neuron populations and not on single neurons.

The field of "NeuroAI" is at the cornerstone of two intrinsic desires of humans: to *create* and to *understand*. The development of new analysis tools and technologies is beneficial to both fields, AI and neuroscience. Will this eventually lead to the development of artificial general intelligence (AGI)? Some predictions imply that AGI is not far into the future [443]. In this context, it is critical to think about how the development and the deployment of AI systems can be safe, ethical, and beneficial to humanity [13, 94]. Even if this important discussion might be more time-pressing than we might think, a key difference between artificial and biological intelligence remains: biological agents can quickly generalize to new environments, while AI systems often remain highly specialized [133]. Besides that, we should be cautious with our expectations from AGI or whole-brain simulations and the insights they might generate. To me, it is not clear that the natural inversion of Richard Feynman’s famous quote “What I cannot create, I do not understand” [123] is necessarily that we can understand everything we build. In my opinion, the more realistic intermediate and beneficial applications of "NeuroAI" seem to be in the clinical area (**Figure 4.1** *top left*). Through a more human-centered approach, statistical frameworks for neuroscience will not only help us understand human perception and learning but also have direct applications in treatments and interventions for patients with neurological disorders.

## 5 | Acknowledgements

I would like to express my deepest gratitude to my parents. I learned everything I know about life from you and you always provided safety, so I could go and explore the world fearlessly. My grandparents for always believing in me. My aunt Simone for being my role model as the first female scientist in our family.

I owe the biggest thank you to my PhD supervisor, Philipp Berens. You always made sure I have the best possible environment for personal and professional growth. I also learned through you how true mentorship looks like. Thank you for your guidance and the incredible opportunities. Thank you, Laura Busse for teaching me how excellent research looks like and for your invaluable feedback. You shaped me into the scientist I am today. I also would like to thank my collaborator Ann. You made working on the project so much more fun.

Dominic, for being my biggest supporter throughout all the years. Thank you for everything. Mila, for being there for me in the darkest times. But also for sharing so many laughter's and sunny days. Cheers to many more! Jesse, for the many scientific discussions and our wonderful friendship. Shout out also to all the strong, smart, and supportive fellow female scientists and friends in Tübingen: Lisa Elina, Kyra, Zina, Hannah, and Auguste. You are a true inspiration and I am so lucky to have you by my side. Some have supported me from the beginning of my PhD, while others I met because of it but they have become just as significant: Torick, I learned so much through our friendship and I would not be where I am today without you. I can't wait for our future adventures.

Last but not least, I would like to thank all the current and previous members of the 'Berens Lab'. Your support and the joyful atmosphere made coming to work a true pleasure.

# Bibliography

- [1] Larry F Abbott. Theoretical neuroscience rising. *Neuron*, 60(3):489–495, 2008.
- [2] Elizabeth D Abercrombie and Barry L Jacobs. Single-unit response of noradrenergic neurons in the locus coeruleus of freely moving cats. i. acutely presented stressful and nonstressful stimuli. *Journal of Neuroscience*, 7(9):2837–2843, 1987.
- [3] WC Abraham and MF Bear. Metaplasticity: the plasticity of synaptic plasticity. *Trends Neurosci*, 19:126–130, 1996.
- [4] George K Aghajanian and C P VanderMaelen. 2-adrenoceptor-mediated hyperpolarization of locus coeruleus neurons: Intracellular studies in vivo. *Science*, 215(4538):1394–1396, 1982.
- [5] George K Aghajanian, J M Cedarbaum, and R Y Wang. Evidence for norepinephrine-mediated collateral inhibition of locus coeruleus neurons. *Brain Research*, 136(3):570–577, 1977.
- [6] Hidenori Aizawa, Megumi Kobayashi, Sayaka Tanaka, Tomoki Fukai, and Hitoshi Okamoto. Molecular characterization of the subnuclei in rat habenula. *Journal of Comparative Neurology*, 520(18):4051–4066, 2012.
- [7] Hidenori Aizawa, Shin Yanagihara, Megumi Kobayashi, Kazue Niisato, Takashi Takekawa, Rie Harukuni, Thomas J McHugh, Tomoki Fukai, Yoshikazu Isomura, and Hitoshi Okamoto. The synchronous activity of lateral habenular neurons is essential for regulating hippocampal theta oscillation. *Journal of Neuroscience*, 33(20):8909–8921, 2013.

- [8] Liz Allen, Jo Scott, Amy Brand, Marjorie Hlava, and Micah Altman. Publishing: Credit where credit is due. *Nature*, 508(7496):312–313, 2014.
- [9] Camille Allene, Joana Lourenço, and Alberto Bacci. The neuronal identity bias behind neocortical gabaergic plasticity. *Trends in Neurosciences*, 38(9):524–534, 2015.
- [10] Jose-Manuel Alonso, W. Martin Usrey, and R. Clay Reid. Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature*, 383(6603):815–819, October 1996.
- [11] Jose-Manuel Alonso, W Martin Usrey, and R Clay Reid. Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. *Journal of Neuroscience*, 21(11):4002–4015, 2001.
- [12] Ryunosuke Amo, Felipe Fredes, Masae Kinoshita, Ryo Aoki, Hidenori Aizawa, Masakazu Agetsuma, et al. The habenulo-raphé serotonergic circuit encodes an aversive expectation value essential for adaptive active avoidance of danger. *Neuron*, 84:1034–1048, 2014.
- [13] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete problems in ai safety. *arXiv preprint arXiv:1606.06565*, 2016.
- [14] David J Anderson. The neural crest cell lineage problem: neuropoiesis? *Neuron*, 3(1):1–12, 1989.
- [15] Alessandra Angelucci and Paul C Bressloff. Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate v1 neurons. *Progress in brain research*, 154:93–120, 2006.
- [16] Detlev Arendt, Jacob M Musser, Clare VH Baker, Aviv Bergman, Connie

- Cepko, Douglas H Erwin, Mihaela Pavlicev, Gerhard Schlosser, Stefanie Widder, Manfred D Laubichler, et al. The origin and evolution of cell types. *Nature Reviews Genetics*, 17(12):744–757, 2016.
- [17] Rubén Armañanzas and Giorgio A Ascoli. Towards the automatic classification of neurons. *Trends in neurosciences*, 38(5):307–318, 2015.
- [18] Gary Aston-Jones and Floyd E Bloom. Activity of norepinephrine-containing locus coeruleus neurons in behaving rats anticipates fluctuations in the sleep-waking cycle. *Journal of Neuroscience*, 1(8):876–886, 1981.
- [19] Gary Aston-Jones and Floyd E Bloom. Norepinephrine-containing locus coeruleus neurons in behaving rats exhibit pronounced responses to non-noxious environmental stimuli. *Journal of Neuroscience*, 1(8):887–900, 1981.
- [20] Gary Aston-Jones and Jonathan D Cohen. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, 28(1):403–450, 2005.
- [21] Gary Aston-Jones, Hiroshi Akaoka, P Charlety, and G Chouvet. Serotonin selectively attenuates glutamate-evoked activation of noradrenergic locus coeruleus neurons. *Journal of Neuroscience*, 11(3):760–769, 1991.
- [22] Gary Aston-Jones, Youhua Zhu, and J P Card. Numerous gabaergic afferents to locus ceruleus in the pericerular dendritic zone: Possible interneuronal pool. *Journal of Neuroscience*, 24(9):2313–2321, 2004.
- [23] Bruno B Averbeck, Peter E Latham, and Alexandre Pouget. Neural correlations, population coding and computation. *Nature reviews neuroscience*, 7(5):358–376, 2006.
- [24] Çağatay Aydın, João Couto, Michele Giugliano, Karl Farrow, and Vincent

- Bonin. Locomotion modulates specific functional cell types in the mouse visual thalamus. *Nature Communications*, 9(1):4882, 2018.
- [25] B. Babadi, A. Casti, Y. Xiao, E. Kaplan, and L. Paninski. A generalized linear model of the impact of direct and indirect inputs to the lateral geniculate nucleus. *Journal of Vision*, 10(10):22–22, August 2010.
- [26] Tom Baden, Philipp Berens, Katrin Franke, Miroslav Roman-Roson, Matthias Bethge, and Thomas Euler. The functional diversity of mouse retinal ganglion cells. *Nature*, 529(7586):1–21, 2016.
- [27] K G Baker, I Törk, J-P Hornung, and P Halasz. The human locus coeruleus complex: An immunohistochemical and three dimensional reconstruction study. *Experimental Brain Research*, 77(2):257–270, 1989.
- [28] Giuseppe Balsamo, Eduardo Blanco-Hernandez, Feng Liang, Robert Konrad Naumann, Stefano Coletta, Andrea Burgalossi, and Patricia Preston-Ferrer. Modular microcircuit organization of the presubicular head-direction map. *Cell Reports*, 39(2), 2022.
- [29] Beatriz Baño-Otálora and Hugh D Piggins. Contributions of the lateral habenula to circadian timekeeping. *Pharmacology Biochemistry and Behavior*, 162:46–54, 2017.
- [30] Yinon M Bar-On, Rob Phillips, and Ron Milo. The biomass distribution on earth. *Proceedings of the National Academy of Sciences*, 115(25):6506–6511, 2018.
- [31] Pouya Bashivan, Kohitij Kar, and James J DiCarlo. Neural population control via deep image synthesis. *Science*, 364(6439):eaav9436, 2019.
- [32] Renata Batista-Brito, Edward Zagha, John M. Ratliff, and Martin Vinck.

- Modulation of cortical circuits by top-down processing and arousal state in health and disease. *Current Opinion in Neurobiology*, 52:172–181, 2018.
- [33] Sébastien Béhuret, Charlotte Deleuze, and Thierry Bal. Corticothalamic synaptic noise as a mechanism for selective attention in thalamic neurons. *Frontiers in Neural Circuits*, 9:80, 2015.
- [34] AL Benabid and L Jeaugey. Cells of the rat lateral habenula respond to high-threshold somatosensory inputs. *Neurosci Lett*, 96:289–294, 1989.
- [35] Corbett Bennett, Sergio Arroyo, and Shaul Hestrin. Subthreshold mechanisms underlying state-dependent modulation of visual responses. *Neuron*, 80(2):350–357, 2013.
- [36] Craig W Berridge and Barry D Waterhouse. The locus coeruleus–noradrenergic system: Modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews*, 42(1):33–84, 2003.
- [37] Kent C. Berridge and Terry E. Robinson. Liking, wanting, and the incentive-sensitization theory of addiction. *American Psychologist*, 71(8):670, 2016.
- [38] Anna Beyeler, Praneeth Namburi, Gordon F Glober, Charlotte Simonnet, Gwendolyn G Calhoon, Gabrielle F Conyers, Robin Luck, Colleen P Wildes, and Kay M Tye. Divergent routing of positive and negative information from the amygdala during memory retrieval. *Neuron*, 90(2):348–361, 2016.
- [39] Tatiana Bezdudnaya, Monica Cano, Yulia Bereshpolova, Carl R. Stoelzel, Jose-Manuel Alonso, and Harvey A. Swadlow. Thalamic burst mode and inattention in the awake LGNd. *Neuron*, 49(3):421–432, February 2006.

- [40] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.
- [41] S Bissière, Y Humeau, and A Lüthi. Dopamine gates ltp induction in lateral amygdala by suppressing feedforward inhibition. *Nat Neurosci*, 6:587–592, 2003.
- [42] Eduardo Blanco-Hernández, Giuseppe Balsamo, Patricia Preston-Ferrer, and Andrea Burgalossi. Sensory and behavioral modulation of thalamic head-direction cells. *Nature Neuroscience*, 27(1):28–33, 2024.
- [43] Henrik Boije, Steffen Rulands, Stefanie Dudczig, Benjamin D Simons, and William A Harris. The independent probabilistic firing of transcription factors: a paradigm for clonal variability in the zebrafish retina. *Developmental cell*, 34(5):532–543, 2015.
- [44] Tom Boissonnet, Matteo Tripodi, and Hiroki Asari. Awake responses suggest inefficient dense coding in the mouse retina. *eLife*, 12, November 2023.
- [45] Gregory Born, Felix A. Schneider-Soupiadis, Sinem Erisken, Agne Vaiceliunaite, Chu Lan Lao, Milad H. Mobarhan, Martin A. Spacek, Gaute T. Einevoll, and Laura Busse. Corticothalamic feedback sculpts visual spatial integration in mouse thalamus. *Nature Neuroscience*, 24(12):1711–1720, November 2021.
- [46] Dante S. Bortone, Shawn R. Olsen, and Massimo Scanziani. Translaminar inhibitory cells recruited by layer 6 corticothalamic neurons suppress visual cortex. *Neuron*, 82(2):474–485, 2014.
- [47] William H Bosking, Ying Zhang, Brett Schofield, and David Fitzpatrick.

- Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of neuroscience*, 17(6):2112–2127, 1997.
- [48] Mihail Bota and Larry W Swanson. The neuron classification problem. *Brain research reviews*, 56(1):79–88, 2007.
- [49] George EP Box and Norman R Draper. *Empirical model-building and response surfaces*. John Wiley & Sons, 1987.
- [50] Vincent Breton-Provencher and Mriganka Sur. Active control of arousal by a locus coeruleus gabaergic circuit. *Nature Neuroscience*, 22(2):218–228, 2019.
- [51] Vincent Breton-Provencher, Geoffrey T Drummond, Jing Feng, Yuchio Li, and Mriganka Sur. Spatiotemporal dynamics of noradrenaline during learned behaviour. *Nature*, 606(7915):732–738, 2022.
- [52] F Briggs and W M Usrey. Corticogeniculate feedback and visual processing in the primate. *Journal of Physiology*, 589(1):33–40, 2010.
- [53] F Briggs and W M Usrey. Corticogeniculate feedback and parallel processing in the primate visual system. *Journal of Physiology*, 589(1):33–40, 2011.
- [54] Farran Briggs and W Martin Usrey. Emerging views of corticothalamic function. *Current opinion in neurobiology*, 18(4):403–407, 2008.
- [55] Farran Briggs and W Martin Usrey. Corticogeniculate feedback and visual processing in the primate. *The Journal of physiology*, 589(1):33–40, 2011.
- [56] Vincent Bringuier, Frederic Chavane, Larry Glaeser, and Yves Frégnac. Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science*, 283(5402):695–699, 1999.

- [57] ES Bromberg-Martin, M Matsumoto, and O Hikosaka. Distinct tonic and phasic anticipatory activity in lateral habenula and dopamine neurons. *Neuron*, 67:144–155, 2010.
- [58] Bundesministerium der Justiz und für Verbraucherschutz. Tierschutzgesetz in der fassung der bekanntmachung vom 18. mai 2006 (bgbl. i s. 1206, 1313), das zuletzt durch artikel 280 der verordnung vom 19. juni 2020 (bgbl. i s. 1328) geändert worden ist. Bundesgesetzblatt, 2006. URL <https://www.gesetze-im-internet.de/tierschg/BJNR012770972.html>. Last amended by Article 280 of the Ordinance of 19 June 2020 (Federal Law Gazette I p. 1328).
- [59] Andrea Buraloss, L Herfst, M von Heimendahl, H Förste, K Haskic, M Schmidt, and Michael Brecht. Microcircuits of functionally identified neurons in the rat medial entorhinal cortex. *Neuron*, 70(4):773–786, 2011.
- [60] Laura Busse, Jessica A Cardin, M Eugenia Chiappe, Michael M Halassa, Matthew J McGinley, Takayuki Yamashita, and Aman B Saleem. Sensation during active behaviors. *Journal of Neuroscience*, 37(45):10826–10834, 2017.
- [61] D. A. Butts, C. Weng, J. Jin, J.-M. Alonso, and L. Paninski. Temporal precision in the visual pathway through the interplay of excitation and stimulus-driven suppression. *Journal of Neuroscience*, 31(31):11313–11327, August 2011.
- [62] Daniel A. Butts. Data-driven approaches to understanding visual neuron activity. *Annual Review of Vision Science*, 5(1):451–477, September 2019.
- [63] Gyorgy Buzsaki. Large-scale recording of neuronal ensembles. *Nature Neuroscience*, 7(5):446–451, 2004.

- [64] S Caldecott-Hazard, J Mazziotta, and M Phelps. Cerebral correlates of depressed behavior in rats, visualized using 14c-2-deoxyglucose autoradiography. *Journal of Neuroscience*, 8(6):1951–1961, 1988.
- [65] Jonathan R Cannon and J Timothy Greenamyre. Neun is not a reliable marker of dopamine neurons in rat substantia nigra. *Neuroscience Letters*, 464(1):14–17, 2009.
- [66] Monica Cano, Tatiana Bezdudnaya, Harvey A Swadlow, and Jose-Manuel Alonso. Brain state and contrast sensitivity in the awake visual thalamus. *Nature Neuroscience*, 9(10):1240–1242, September 2006.
- [67] Rosa Cao and Daniel Yamins. Explanatory models in neuroscience, part 1: Taking mechanistic abstraction seriously. *Cognitive Systems Research*, page 101244, 2024.
- [68] Matteo Carandini, David J Heeger, and J Anthony Movshon. Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*, 17(21):8621–8644, 1997.
- [69] Matthew E Carter, Ofer Yizhar, Sachiko Chikahisa, Hien Nguyen, Antoine Adamantidis, Seiji Nishino, Karl Deisseroth, and Luis de Lecea. Tuning arousal with optogenetic modulation of locus coeruleus neurons. *Nature Neuroscience*, 13(12):1526–1533, 2010.
- [70] John M Cedarbaum and George K Aghajanian. Activation of locus coeruleus neurons by peripheral stimuli: Modulation by a collateral inhibitory mechanism. *Life Sciences*, 23(13):1383–1392, 1978.
- [71] Daniel J Chandler, Wen-Jun Gao, and Barry D Waterhouse. Heterogeneous organization of the locus coeruleus projections to prefrontal and motor cor-

- tices. *Proceedings of the National Academy of Sciences*, 111(18):6816–6821, 2014.
- [72] Daniel J Chandler, Phillip Jensen, Jordan G McCall, Ashley E Pickering, Lindsay A Schwarz, and Nelson K B Totah. Redefining noradrenergic neuromodulation of behavior: Impacts of a modular locus coeruleus architecture. *Journal of Neuroscience*, 39(42):8239–8249, 2019.
- [73] Fu-Jen Chen and Susan J Sara. Locus coeruleus activation by foot shock or electrical stimulation inhibits amygdala neurons. *Neuroscience*, 144(2):472–481, 2007.
- [74] Clement Chiang and Gary Aston-Jones. Response of locus coeruleus neurons to footshock stimulation is mediated by neurons in the rostral ventral medulla. *Neuroscience*, 53(3):705–715, 1993.
- [75] EJ Chichilnisky. A simple white noise analysis of neuronal light responses. *Network: computation in neural systems*, 12(2):199, 2001.
- [76] Greg R Christoph, Robert J Leonzio, and Karen S Wilcox. Stimulation of the lateral habenula inhibits dopamine-containing neurons in the substantia nigra and ventral tegmental area of the rat. *Journal of Neuroscience*, 6(3):613–619, 1986.
- [77] Patricia Smith Churchland and Terrence Joseph Sejnowski. *The computational brain*. MIT press, 1992.
- [78] Liset Cintra, Soledad Díaz-Cintra, Thomas Kemper, and Peter J Morgane. Nucleus locus coeruleus: A morphometric golgi study in rats of three age groups. *Brain Research*, 247(1):17–28, 1982.
- [79] Francisco Clascá, Pablo Rubio-Garrido, and Denis Jabaudon. Unveiling the

- diversity of thalamocortical neuron subtypes. *European Journal of Neuroscience*, 35(10):1524–1532, 2012.
- [80] Roger L Clem and Richard L Huganir. Calcium-permeable ampa receptor dynamics mediate fear memory erasure. *science*, 330(6007):1108–1112, 2010.
- [81] Ruben Coen-Cagli, Adam Kohn, and Odelia Schwartz. Flexible gating of contextual influences in natural vision. *Nature Neuroscience*, 18(11):1648–1655, October 2015.
- [82] Jeremiah Y Cohen, Sebastian Haesler, Linh Vong, Bradford B Lowell, and Naoshige Uchida. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *nature*, 482(7383):85–88, 2012.
- [83] Mauro Congiu, Massimo Trusel, Marco Pistis, Manuel Mameli, and Salvatore Lecca. Opposite responses to aversive stimuli in lateral habenula neurons. *European Journal of Neuroscience*, 50(6):2921–2930, 2019.
- [84] Mauro Congiu, Sarah Mondoloni, Ioannis S Zouridis, Lisa Schmors, Salvatore Lecca, Arnaud L Lalive, Kyllian Ginggen, Fei Deng, Philipp Berens, Rosa Chiara Paolicelli, et al. Plasticity of neuronal dynamics in the lateral habenula for cue-punishment associative learning. *Molecular Psychiatry*, pages 1–10, 2023.
- [85] Neil L Corteen, Tim M Cole, Ankur Sarna, Werner Sieghart, and J Douglas Swinny. Localization of gaba-a receptor alpha subunits on neurochemically distinct cell types in the rat locus coeruleus. *European Journal of Neuroscience*, 34(2):250–262, 2011.
- [86] Council of the European Communities. Council directive 86/609/eec of 24 november 1986 on the approximation of laws, regulations and administra-

- tive provisions of the member states regarding the protection of animals used for experimental and other scientific purposes. Official Journal of the European Communities, 1986. URL <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:31986L0609>. OJ L 358, 18.12.1986, p. 1–28.
- [87] Benjamin R Cowley, Patricia L Stan, Jonathan W Pillow, and Matthew A Smith. Compact deep neural network models of visual cortex. *bioRxiv*, 2023.
- [88] Shane R Crandall, Scott J Cruikshank, and Barry W Connors. A corticothalamic switch: controlling the thalamus with dynamic synapses. *Neuron*, 86(3):768–782, 2015.
- [89] Davide Crombie, Martin A. Spacek, Christian Leibold, and Laura Busse. Spiking activity in the visual thalamus is coupled to pupil dynamics across temporal scales. *PLoS Biology*, 22(5):e3002614, May 2024.
- [90] Yihui Cui, Yan Yang, Zheyi Ni, Yiyan Dong, Guohong Cai, Alexandre Foncelles, Shuangshuang Ma, Kangning Sang, Siyang Tang, Yuezhou Li, et al. Astroglial kir4. 1 in the lateral habenula drives neuronal bursts in depression. *Nature*, 554(7692):323–327, 2018.
- [91] Yihui Cui, Shaohua Hu, and Hailan Hu. Lateral habenular burst firing as a target of the rapid antidepressant effects of ketamine. *Trends in Neurosciences*, 42(3):179–191, 2019.
- [92] John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, 17(11):1500–1509, 2014.
- [93] András Czurkó, John Huxter, Ying Li, Balázs Hangya, and Robert U Muller. Theta phase classification of interneurons in the hippocampal for-

- mation of freely moving rats. *Journal of Neuroscience*, 31(8):2938–2947, 2011.
- [94] David Dalrymple, Joar Skalse, Yoshua Bengio, Stuart Russell, Max Tegmark, Sanjit Seshia, Steve Omohundro, Christian Szegedy, Ben Goldhaber, Nora Ammann, et al. Towards guaranteed safe ai: A framework for ensuring robust and reliable ai systems. *arXiv preprint arXiv:2405.06624*, 2024.
- [95] Michael Davis, David L Walker, Leigh Miles, and Christian Grillon. Phasic vs sustained fear in rats and humans: role of the extended amygdala in fear vs anxiety. *Neuropsychopharmacology*, 35(1):105–135, 2010.
- [96] Peter Dayan and Laurence F Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT press, 2005.
- [97] Bradley Dearnley, Melissa Jones, Martynas Dervinis, and Michael Okun. Brain state transitions primarily impact the spontaneous rate of slow-firing neurons. *Cell Reports*, 42(10):113185, October 2023.
- [98] Nathalie Dehorter, Gabriele Ciceri, Giorgia Bartolini, Lynette Lim, Isabel Del Pino, and Oscar Marín. Tuning of fast-spiking interneuron properties by an activity-dependent transcriptional switch. *Science*, 349(6253):1216–1220, 2015.
- [99] Michael Deistler, Kyra L. Kadhim, Ziwei Huang, Cornelius Schröder, Janne Lappalainen, Jonas Beck, Philipp Berens, Pedro Goncalves, and Jakob Macke. Training networks of biophysically detailed biophysical neurons with thousands of parameters. In *Proceedings of the Computational and Systems Neuroscience Conference (COSYNE)*, 2024. Poster presented at the COSYNE 2024 Conference.

- [100] Onkar S Dhande and Andrew D Huberman. Retinal ganglion cell maps in the brain: implications for visual processing. *Current opinion in neurobiology*, 24:133–142, 2014.
- [101] Onkar S Dhande, Benjamin K Stafford, Jung-Hwan A Lim, and Andrew D Huberman. Contributions of retinal ganglion cells to subcortical visual processing and behaviors. *Annual review of vision science*, 1:291–328, 2015.
- [102] Maria Diamantaki, Markus Frey, Philipp Berens, Patricia Preston-Ferrer, and Andrea Burgalossi. Sparse activity of identified dentate granule cells during spatial exploration. *Elife*, 5:e20252, 2016.
- [103] Maria Diamantaki, Silvia Coletta, Karim Nasr, Roozbeh Zeraati, Swantje Lathunus, Philipp Berens, Pablo Preston-Ferrer, and Andrea Burgalossi. Manipulating hippocampal place cell activity by single-cell stimulation in freely moving mice. *Cell Reports*, 23(1):32–38, 2018.
- [104] James J DiCarlo, Davide Zoccolan, and Nicole C Rust. How does the brain solve visual object recognition? *Neuron*, 73(3):415–434, 2012.
- [105] Mathias Dietz and Go Ashida. Computational models of binaural processing. *Binaural Hearing: With 93 Illustrations*, pages 281–315, 2021.
- [106] Lu Ding, Gianluca Balsamo, Han Chen, Elias Blanco-Hernandez, Ioannis S Zouridis, Rainer Naumann, Pablo Preston-Ferrer, and Andrea Burgalossi. Juxtacellular opto-tagging of hippocampal cal neurons in freely moving mice. *eLife*, 11:e71720, 2022.
- [107] Annette J Dobson and Adrian G Barnett. *An introduction to generalized linear models*. Chapman and Hall/CRC, 2018.
- [108] Adrien Doerig, Rowan P Sommers, Katja Seeliger, Blake Richards, Jenann

- Ismael, Grace W Lindsay, Konrad P Kording, Talia Konkle, Marcel AJ Van Gerven, Nikolaus Kriegeskorte, et al. The neuroconnectionist research programme. *Nature Reviews Neuroscience*, 24(7):431–450, 2023.
- [109] WQ Dong, OB Wilson, MH Skolnick, and N Dafny. Hypothalamic, dorsal raphe and external electrical stimulation modulate noxious evoked responses of habenula neurons. *Neuroscience*, 48:933–940, 1992.
- [110] Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017. URL <https://arxiv.org/abs/1702.08608>.
- [111] URSULA C Drager and DH Hubel. Topography of visual and somatosensory projections to mouse superior colliculus. *Journal of Neurophysiology*, 39(1):91–101, 1976.
- [112] Séverine Durand, Ramakrishnan Iyer, Kenji Mizuseki, Saskia De Vries, Stefan Mihalas, and R. Clay Reid. A comparison of visual response properties in the lateral geniculate nucleus and primary visual cortex of awake and anesthetized mice. *Journal of Neuroscience*, 36(48):12144–12156, 2016.
- [113] E Edelman and V Lessmann. Dopamine modulates spike timing-dependent plasticity and action potential properties in ca1 pyramidal neurons of acute rat hippocampal slices. *Front Synaptic Neurosci*, 3:6, 2011.
- [114] Matthew Ennis and Gary Aston-Jones. Evidence for self- and neighbor-mediated postactivation inhibition of locus coeruleus neurons. *Brain Research*, 374(2):299–305, 1986.
- [115] Sinem Erisken, Agne Vaiceliunaite, Ovidiu Jurjut, Matilde Fiorini, Steffen Katzner, and Laura Busse. Effects of locomotion extend throughout the mouse early visual system. *Current Biology*, 24(24):2899–2907, 2014.

- [116] Sinem Erisken, Agne Vaiceliunaite, Ovidiu Jurjut, Matilde Fiorini, Steffen Katzner, and Laura Busse. Effects of locomotion extend throughout the mouse early visual system. *Current Biology*, 24(24):2899–2907, 2014.
- [117] Oxana Eschenko. The role of the locus coeruleus in cellular and systems memory consolidation. In Denise Manahan-Vaughan, editor, *Handbook of Behavioral Neuroscience*, volume 28, pages 327–347. Elsevier, 2018.
- [118] A Aldo Faisal, Luc PJ Selen, and Daniel M Wolpert. Noise in the nervous system. *Nature reviews neuroscience*, 9(4):292–303, 2008.
- [119] Marc Fakhoury. The habenula in psychiatric disorders: more than three decades of translational investigation. *Neuroscience & Biobehavioral Reviews*, 83:721–735, 2017.
- [120] Yizhou Fan, Peng Chen, Yuchio Li, Kevin Cui, Diana M Noel, Edward D Cummins, Darrell J Peterson, Richard W Brown, and Ming Zhu. Corticosterone administration up-regulated expression of norepinephrine transporter and dopamine -hydroxylase in rat locus coeruleus and its terminal regions. *Journal of Neurochemistry*, 128(3):445–458, 2014.
- [121] Gidon Felsen and Yang Dan. A natural approach to studying vision. *Nature Neuroscience*, 8(12):1643–1646, 2005.
- [122] F. Feng, P. Samarth, D. Paré, and S. S. Nair. Mechanisms underlying the formation of the amygdalar fear memory trace: A computational perspective. *Neuroscience*, 322:370–376, 2016.
- [123] Richard P. Feynman. Richard feynman quotes, 1988. URL [https://simple.wikiquote.org/wiki/Richard\\_Feynman](https://simple.wikiquote.org/wiki/Richard_Feynman). Quote found on Feynman’s blackboard at the time of his death in 1988.

- [124] Gord Fishell and Nathaniel Heintz. The neuron identity problem: form meets function. *Neuron*, 80(3):602–612, 2013.
- [125] Meaghan E Flanigan, Hossein Aleyasin, Lin Li, Colin J Burnett, Kevin L Chan, Katherine B LeClair, et al. Orexin signaling in gabaergic lateral habenula neurons modulates aggressive behavior in male mice. *Nat Neurosci*, 23:638–650, 2020.
- [126] Chris Fraley and Adrian E Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, 97(458):611–631, 2002.
- [127] Jaclyn Ellen Frandolig, Chanel Joylae Matney, Kihwan Lee, Juhyun Kim, Maxime Chevée, Su-Jeong Kim, Aaron Andrew Bickert, and Solange Pezon Brown. The synaptic organization of layer 6 circuits reveals inhibition as a major output of a neocortical sublamina. *Cell Reports*, 28(12):3131–3143, 2019.
- [128] Felix Franke, Robert Pröpper, Henrik Alle, Philipp Meier, Jörg RP Geiger, Klaus Obermayer, and Matthias HJ Munk. Spike sorting of synchronous spikes from local neuron ensembles. *Journal of Neurophysiology*, 114(4):2535–2549, 2015.
- [129] Katrin Franke, Konstantin F Willeke, Kayla Ponder, Mario Galdamez, Na Zhou, Taliah Muhammad, Saumil Patel, Emmanouil Froudarakis, Jacob Reimer, Fabian H Sinz, et al. State-dependent pupil dilation rapidly shifts visual feature selectivity. *Nature*, 610(7930):128–134, 2022.
- [130] Juan A Gallego, Matthew G Perich, Lee E Miller, and Sara A Solla. Neural manifolds for the control of movement. *Neuron*, 94(5):978–984, 2017.
- [131] Sacha Garcia, Alessio Paolo Buccino, and Pierre Yger. How do spike col-

- lisions affect spike sorting performance? *eNeuro*, 9(5):ENEURO.0105–22.2022, 2022.
- [132] Silvia Gasparini, Jason M Resch, Amanda M Gore, Laura Peltekian, and Joel C Geerling. Pre-locus coeruleus neurons in rat and mouse. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 320(3):R342–R361, 2021.
- [133] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*, 2022. URL <https://arxiv.org/abs/1811.12231>.
- [134] Wulfram Gerstner, Henning Sprekeler, and Gustavo Deco. Theory and simulation in neuroscience. *science*, 338(6103):60–65, 2012.
- [135] Joshua I Glaser, Ari S Benjamin, Roozbeh Farhoodi, and Konrad P Kording. The roles of supervised machine learning in systems neuroscience. *Progress in neurobiology*, 175:126–137, 2019.
- [136] C Gold, DA Henze, Christof Koch, and Gyorgy Buzsaki. On the origin of the extracellular action potential waveform: A modeling study. *Journal of Neurophysiology*, 95(5):3113–3128, 2006.
- [137] Matías A Goldin, Baptiste Lefebvre, Samuele Virgili, Mathieu Kim Pham Van Cang, Alexander Ecker, Thierry Mora, Ulisse Ferrari, and Olivier Marre. Context-dependent selectivity to natural images in the retina. *Nature Communications*, 13(1):5556, 2022.
- [138] Pieter M. Goltstein, Sandra Reinert, Tobias Bonhoeffer, and Mark Hübener. Mouse visual cortex areas represent perceptual and semantic features

of learned visual categories. *Nature Neuroscience*, 24(10):1441–1451, September 2021.

- [139] Rui Gong, Shan Xu, Ann M Hermundstad, Yan Yu, and Scott M Sternson. Hindbrain double-negative feedback mediates palatability-guided food and water consumption. *Cell*, 182(6):1589–1605.e22, 2020.
- [140] Shiaoqing Gong, Martin Doughty, Carroll R Harbaugh, Alexander Cummins, Mary E Hatten, Nathaniel Heintz, and Charles R Gerfen. Targeting cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *Journal of Neuroscience*, 27(37):9817–9823, sep 2007.
- [141] David GoodSmith, Xiang Chen, Cheng Wang, Sang-Hun Kim, Hongjun Song, Andrea Burgalossi, Kimberly M Christian, and James J Knierim. Spatial representations of granule cells and mossy cells of the dentate gyrus. *Neuron*, 93(3):677–690.e5, 2017.
- [142] Elena G Govorunova, Oleg A Sineshchekov, Roger Janz, Xiaoqin Liu, and John L Spudich. Natural light-gated anion channels: A family of microbial rhodopsins for advanced optogenetics. *Science*, 349(6248):647–650, 2015.
- [143] Austin R. Graves, Shannon J. Moore, Erik B. Bloss, Brett D. Mensh, William L. Kath, and Nelson Spruston. Hippocampal pyramidal neurons comprise two distinct cell types that are countermodulated by metabotropic receptors. *Neuron*, 76(4):776–789, 2012.
- [144] Benjamin F Grewe, Jan Gründemann, Lacey J Kitch, Jerome A Lecoq, Josh G Parker, Jesse D Marshall, et al. Neural ensemble dynamics underlying a long-term associative memory. *Nature*, 543:670–675, 2017.
- [145] Matthew S Grubb and Ian D Thompson. Quantitative characterization of

- visual response properties in the mouse dorsal lateral geniculate nucleus. *Journal of Neurophysiology*, 90(6):3594–3607, 2003.
- [146] Reinhard Grzanna and ME Molliver. The locus coeruleus in the rat: An immunohistochemical delineation. *Neuroscience*, 5(1):21–40, 1980.
- [147] William Guido. Development, form, and function of the mouse visual thalamus. *Journal of Neurophysiology*, 120(1):211–225, July 2018.
- [148] Carolyn W Harley. Norepinephrine and the dentate gyrus. *Progress in brain research*, 163:299–318, 2007.
- [149] Kenneth D. Harris. A shift test for independence in generic time series. *arXiv*, pages 1–6, 2020.
- [150] Yoshiko Hashikawa, Koichi Hashikawa, Mark A Rossi, Marcus L Basiri, Yuejia Liu, Nathan L Johnston, Omar R Ahmad, and Garret D Stuber. Transcriptional and spatial resolution of cell types in the mammalian habenula. *Neuron*, 106(5):743–758, 2020.
- [151] Demis Hassabis, Dhharshan Kumaran, Christopher Summerfield, and Matthew Botvinick. Neuroscience-inspired artificial intelligence. *Neuron*, 95(2):245–258, 2017.
- [152] J. Michael Hasse and Farran Briggs. Corticogeniculate feedback sharpens the temporal precision and spatial resolution of visual signals in the ferret. *Proceedings of the National Academy of Sciences of the United States of America*, 114(30):E6222–E6230, 2017.
- [153] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

- [154] Darrell A Henze, Zsolt Borhegyi, Jozsef Csicsvari, Airi Mamiya, Kenneth D Harris, and Gyorgy Buzsaki. Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *Journal of Neurophysiology*, 84(1):390–400, 2000.
- [155] Laurens Herfst, Andrea Buralossio, Kenan Haskic, Jan J Tukker, Martin Schmidt, and Michael Brecht. Friction-based stabilization of juxtacellular recordings in freely moving rats. *Journal of Neurophysiology*, 108(2):697–707, 2012.
- [156] Miles Herkenham and Walle JH Nauta. Efferent connections of the habenular nuclei in the rat. *Journal of Comparative Neurology*, 187(1):19–47, 1979.
- [157] Jose L Herrero, Marc A Gieselmann, Mehdi Sanayei, and Alexander Thiele. Attention-induced variance and noise correlation reduction in macaque v1 is mediated by nmda receptors. *Neuron*, 78(4):729–739, 2013.
- [158] Cyril Herry and Joshua P Johansen. Encoding of fear learning and memory in distributed neuronal circuits. *Nat Neurosci*, 17:1644–1654, 2014.
- [159] Lisa Hickey, Yulan Li, Sarah J Fyson, Thomas C Watson, Rachael Perrins, Jack Hewinson, Anja G Teschemacher, Hidemasa Furue, Bridget M Lumb, and Ashley E Pickering. Optoactivation of locus ceruleus neurons evokes bidirectional changes in thermal nociception in rats. *Journal of Neuroscience*, 34(12):4148–4160, 2014.
- [160] Okihide Hikosaka. The habenula: from stress evasion to value-based decision-making. *Nature reviews neuroscience*, 11(7):503–513, 2010.
- [161] GE Hinton and S Roweis. Stochastic neighbor embedding. *Advances in neural information processing systems*, 15:857–864, 2002.

- [162] Akio Hirata, Juan Aguilar, and Manuel A Castro-Alamancos. Noradren-  
ergic activation amplifies bottom-up and top-down signal-to-noise ratios in  
sensory thalamus. *Journal of Neuroscience*, 26(16):4426–4436, 2006.
- [163] Hajime Hirata and Gary Aston-Jones. A novel long-latency response of  
locus coeruleus neurons to noxious stimuli: Mediation by peripheral c-  
fibers. *Journal of Neurophysiology*, 71(5):1752–1761, 1994.
- [164] Judith A Hirsch, Xin Wang, Friedrich T Sommer, and Luis M Martinez.  
How inhibitory circuits in the thalamus serve vision. *Annual Review of  
Neuroscience*, 38:309–329, 2015.
- [165] Scott Hirschberg, Yulan Li, Amy Randall, Eric J Kremer, and Ashley E  
Pickering. Functional dichotomy in spinal- vs prefrontal-projecting locus  
coeruleus modules splits descending noradrenergic analgesia from ascend-  
ing aversion and anxiety in rats. *eLife*, 6:e29808, 2017.
- [166] Alan L Hodgkin and Andrew F Huxley. A quantitative description of mem-  
brane current and its application to conduction and excitation in nerve. *The  
Journal of physiology*, 117(4):500, 1952.
- [167] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estima-  
tion for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.
- [168] Robert N Holdefer and Barry L Jacobs. Phasic stimulation of the locus  
coeruleus: effects on activity in the lateral geniculate nucleus. *Experimen-  
tal brain research*, 79:444–452, 1994.
- [169] Astrid R Houweling and Michael Brecht. Behavioural report of single neu-  
ron stimulation in somatosensory cortex. *Nature*, 451(7174):65–68, 2008.
- [170] Astrid R Houweling, Guy Doron, Beat C Voigt, Laurens J Herfst, and  
Michael Brecht. Nanostimulation: Manipulation of single neuron activ-

- ity by juxtacellular current injection. *Journal of Neurophysiology*, 103(3): 1696–1704, 2010.
- [171] W. M. Howe and P. J. Kenny. Burst firing sets the stage for depression. *Nature*, 554:304–305, 2018.
- [172] Hailan Hu, Yihui Cui, and Yan Yang. Circuits and functions of the lateral habenula in health and in disease. *Nature Reviews Neuroscience*, 21(5): 277–295, 2020.
- [173] L Huang, Y Xi, Y Peng, Y Yang, X Huang, Y Fu, et al. A visual circuit related to habenula underlies the antidepressive effects of light therapy. *Neuron*, 102:128–142.e8, 2019.
- [174] Ziwei Huang, Yanli Ran, Jonathan Oesterle, Thomas Euler, and Philipp Berens. Estimating smooth and sparse neural receptive fields with a flexible spline basis. *Neurons, Behavior, Data analysis, and Theory*, 8:1–14, 2021.
- [175] Donald R Humphrey, EM Schmidt, and WD Thompson. Predicting measures of motor performance from multiple cortical spike trains. *Science*, 170(3959):758–762, 1970.
- [176] Jennifer M. Ichida and Vivien A. Casagrande. Organization of the feedback pathway from striate cortex (V1) to the lateral geniculate nucleus (LGN) in the owl monkey (*aotus trivirgatus*). *Journal of Comparative Neurology*, 454(3):272–283, November 2002.
- [177] Hiroshi Ito, Anne C Sales, Christopher H Fry, Azusa J Kanai, Marcus J Drake, and Ashley E Pickering. Probabilistic, spinally-gated control of bladder pressure and autonomous micturition by barrington’s nucleus crh neurons. *eLife*, 9:e56605, 2020.
- [178] MH James, JL Charnley, JR Flynn, DW Smith, and CV Dayas. Propensity

- to 'relapse' following exposure to cocaine cues is associated with the recruitment of specific thalamic and epithalamic nuclei. *Neuroscience*, 199: 235–242, 2011.
- [179] Mehrdad Jazayeri and J Anthony Movshon. Optimal representation of sensory information by neural populations. *Nature neuroscience*, 9(5):690–696, 2006.
- [180] Thomas C Jhou, Howard L Fields, Mark G Baxter, Clifford B Saper, and Peter C Holland. The rostromedial tegmental nucleus (rmtg), a gabaergic afferent to midbrain dopamine neurons, encodes aversive stimuli and inhibits motor responses. *Neuron*, 61(5):786–800, 2009.
- [181] Huifang Ji and Paul D Shepard. Lateral habenula stimulation inhibits rat midbrain dopamine neurons through a gabaa receptor-mediated mechanism. *Journal of Neuroscience*, 27(26):6923–6930, 2007.
- [182] Xiang Jin, Shu Li, Brian Bondy, Weilin Zhong, Matthew F Oginsky, Yong Wu, Clay M Johnson, Shu Zhang, Ningren Cui, and Changyu Jiang. Identification of a group of gabaergic neurons in the dorsomedial area of the locus coeruleus. *PLoS ONE*, 11(1):e0146470, 2016.
- [183] M Jing, Y Li, J Zeng, P Huang, M Skirzewski, O Kljakic, et al. An optimized acetylcholine sensor for monitoring in vivo cholinergic activity. *Nat Methods*, 17:1139–1146, 2020.
- [184] Helen E. Jones, Ian M. Andolina, Bashir Ahmed, Stewart D. Shipp, Jake T. C. Clements, Kenneth L. Grieve, Javier Cudeiro, Thomas E. Salt, and Adam M. Sillito. Differential feedback modulation of center and surround mechanisms in parvocellular cells in the visual thalamus. *The Journal of Neuroscience*, 32(45):15946–15951, November 2012.

- [185] Ilenna Simone Jones and Konrad Paul Kording. Efficient optimization of ode neuron models using gradient descent. *arXiv preprint arXiv:2407.04025*, 2024.
- [186] Judson P Jones and Larry A Palmer. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of neurophysiology*, 58(6):1187–1211, 1987.
- [187] Z. Josh Huang, Hongkui Zeng, Z Josh Huang, and Hongkui Zeng. Genetic approaches to neural circuits in the mouse. *Annual review of neuroscience*, 36(1):183–215, 2013.
- [188] David Jung, Sungkwang Kim, Ahmed Sariev, Farzana Sharif, Dongsu Kim, and Sebastien Royer. Dentate granule and mossy cells exhibit distinct spatiotemporal responses to local change in a one-dimensional landscape of visual-tactile cues. *Scientific Reports*, 9(1):9545, 2019.
- [189] Sergey Kalinin, Douglas L Feinstein, Hui-Ling Xu, Giselle Huesa, Dale A Pelligrino, and Edgardo Galea. Degeneration of noradrenergic fibres from the locus coeruleus causes tight-junction disorganisation in the rat brain. *European Journal of Neuroscience*, 24:3393–3400, 2006.
- [190] Rishi M Kalwani, Sandeep Joshi, and Joshua I Gold. Phasic activation of individual neurons in the locus ceruleus/subceruleus complex of monkeys reflects rewarded decisions to go but not stop. *Journal of Neuroscience*, 34(41):13656–13669, 2014.
- [191] Seyed-Mahdi Khaligh-Razavi and Nikolaus Kriegeskorte. Deep supervised, but not unsupervised, models may explain it cortical representation. *PLoS computational biology*, 10(11):e1003915, 2014.
- [192] Tim C Kietzmann, Courtney J Spoerer, Lynn KA Sörensen, Radoslaw M

- Cichy, Olaf Hauk, and Nikolaus Kriegeskorte. Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences*, 116(43):21854–21863, 2019.
- [193] J Kim, C J Matney, A Blankenship, S Hestrin, and S P Brown. Layer 6 corticothalamic neurons activate a cortical output layer, layer 5a. *J Neurosci*, 34(29):9656–9664, 2014.
- [194] U. Kim. Topographic commissural and descending projections of the habenula in the rat. *Journal of Comparative Neurology*, 513(2):173–187, 2009.
- [195] U. Kim and S. Y. Chang. Dendritic morphology, local circuitry, and intrinsic electrophysiology of neurons in the rat medial and lateral habenular nuclei of the epithalamus. *Journal of Comparative Neurology*, 483(2):236–250, 2005.
- [196] Bálint Király and Balázs Hangya. Navigating the statistical minefield of model selection and clustering in neuroscience. *Eneuro*, 9(4), 2022.
- [197] Megan A Kirchgeßner, Alexis D Franklin, and Edward M Callaway. Distinct “driving” versus “modulatory” influences of different visual corticothalamic pathways. *Current Biology*, 31(23):5121–5137, 2021.
- [198] Thomas Klausberger, Peter J Magill, Laszlo F Marton, J Douglas B Roberts, Paula M Cobden, Gyorgy Buzsaki, and Peter Somogyi. Brain-state- and cell-type-specific firing of hippocampal interneurons in vivo. *Nature*, 421(6925):844–848, 2003.
- [199] David Klindt, Alexander S Ecker, Thomas Euler, and Matthias Bethge. Neural system identification for large populations separating “what” and “where”. *Advances in neural information processing systems*, 30, 2017.

- [200] Dmitry Kobak and Philipp Berens. The art of using t-sne for single-cell transcriptomics. *Nature communications*, 10(1):5416, 2019.
- [201] Bryan Kolb and Ian Q Whishaw. Brain plasticity and behavior. *Annual review of psychology*, 49(1):43–64, 1998.
- [202] A. B. Kowski, R. W. Veh, and T. Weiss. Dopaminergic activation excites rat lateral habenular neurons in vivo. *Neuroscience*, 161(4):1154–1165, 2009.
- [203] Yuji Koyama, Eiichi Jodo, and Yutaka Kayama. Sensory responsiveness of “broad-spike” neurons in the laterodorsal tegmental nucleus, locus coeruleus and dorsal raphe of awake rats: Implications for cholinergic and monoaminergic neuron-specific responses. *Neuroscience*, 63(4):1021–1031, 1994.
- [204] Rüdiger Krahe and Fabrizio Gabbiani. Burst firing in sensory systems. *Nature Reviews Neuroscience*, 5(1):13–23, 2004.
- [205] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.
- [206] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.
- [207] Ching-Wei Kuo, Jen-Jarlin Hsieh, Huei-Chen Tsai, Yi-Ming Kuo, He-Yuan Yau, Chih-Cheng Chen, Ren-Yuan Chen, Han-Ching Yang, and Ming-Kai Min. Inhibitory interneurons regulate phasic activity of noradrenergic neurons in the mouse locus coeruleus and functional implications. *Journal of Physiology*, 598(18):4003–4029, 2020.
- [208] AL Lalive, M Congiu, C Lewis, D Groos, JA Clerke, A Tchenio, et al.

- Synaptic inhibition in the lateral habenula shapes reward anticipation. *Curr Biol*, 32:1829–1836.e4, 2022.
- [209] Victor AF Lamme and Pieter R Roelfsema. The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23(11):571–579, 2000.
- [210] Stefan Lang, Vincent J Dercksen, Bert Sakmann, and Marcel Oberlaender. Simulation of signal flow in 3d reconstructions of an anatomically realistic neural network in rat vibrissal cortex. *Neural Networks*, 24(9):998–1011, 2011.
- [211] Rebecca P Lawson, Ben Seymour, Eleanor Loh, Antoine Lutti, Raymond J Dolan, Peter Dayan, et al. The habenula encodes negative motivational value associated with primary punishment in humans. *Proc Natl Acad Sci*, 111:11858–11863, 2014.
- [212] Ioannis Lazaridis, O Tzortzi, M Weglage, A Martin, Y Xuan, M Parent, et al. A hypothalamus-habenula circuit controls aversion. *Mol Psychiatry*, 24:1351–1368, 2019.
- [213] Salvatore Lecca, Frank J Meye, and Manuel Mameli. The lateral habenula in addiction and depression: an anatomical, synaptic and behavioral overview. *Eur J Neurosci*, 39:1170–1178, 2014.
- [214] Salvatore Lecca, Assunta Pelosi, Anna Tchenio, Imane Moutkine, Rafael Lujan, Denis Hervé, and Manuel Mameli. Rescue of gabab and girk function in the lateral habenula by protein phosphatase 2a inhibition ameliorates depression-like phenotypes in mice. *Nature medicine*, 22(3):254–261, 2016.
- [215] Salvatore Lecca, Frank Julius Meye, Massimo Trusel, Anna Tchenio, Ju-

- lia Harris, Martin Karl Schwarz, Denis Burdakov, Francois Georges, and Manuel Mameli. Aversive stimuli drive hypothalamus-to-habenula excitation to promote escape behavior. *Elife*, 6:e30697, 2017.
- [216] Joseph LeDoux. The amygdala. *Current biology*, 17(20):R868–R874, 2007.
- [217] Joseph E LeDoux. Emotion circuits in the brain. *Annual review of neuroscience*, 23(1):155–184, 2000.
- [218] D Lee, JH Hyun, K Jung, P Hannan, and HB Kwon. A calcium- and light-gated switch to induce gene expression in activated neurons. *Nat Biotechnol*, 35:858–863, 2017.
- [219] Seung-Hee Lee and Yang Dan. Neuromodulation of brain states. *neuron*, 76(1):209–222, 2012.
- [220] Nicholas A Lesica, Chong Weng, Jianzhong Jin, Chun-I Yeh, Jose-Manuel Alonso, and Garrett B Stanley. Dynamic encoding of natural luminance sequences by lgn bursts. *PLoS Biology*, 4(7):e209, June 2006.
- [221] Nicholas A Lesica, Jianzhong Jin, Chong Weng, Chun-i Yeh, Daniel A Butts, Garrett B Stanley, and Jose-manuel Alonso. Adaptation to stimulus contrast and correlations during natural visual stimulation. *Neuron*, 55(3):479–491, 2007.
- [222] MR Levinstein, DJ Bergkamp, ZK Lewis, A Tsobanoudis, K Hashikawa, GD Stuber, et al. Pacap-expressing neurons in the lateral habenula diminish negative emotional valence. *Genes Brain Behav*, 21:e12801, 2022.
- [223] Michael S Lewicki. A review of methods for spike sorting: The detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 9(4):R53–R78, 1998.

- [224] Bo Li, Joaquin Piriz, Martine Mirrione, Christine Chung, Christophe D Proulx, Daniela Schulz, et al. Synaptic potentiation onto habenula neurons in the learned helplessness model of depression. *Nature*, 470:535–539, 2011.
- [225] H Li, D Pullmann, and TC Jhou. Valence-encoding in the lateral habenula arises from the entopeduncular region. *Elife*, 8:e41223, 2019.
- [226] Kun Li, Tao Zhou, Lujian Liao, Zhongfei Yang, Catherine Wong, Fritz Henn, Roberto Malinow, John R Yates III, and Hailan Hu.  $\beta$ camkii in lateral habenula mediates core symptoms of depression. *science*, 341(6149): 1016–1020, 2013.
- [227] Wen Li, James D Howard, Todd B Parrish, and Jay A Gottfried. Aversive learning enhances perceptual and cortical discrimination of indiscriminable odor cues. *Science*, 319(5871):1842–1845, 2008.
- [228] Yulan Li, Lisa Hickey, Rachael Perrins, Esmeralda Werlen, Anjali A Patel, Scott Hirschberg, Matthew W Jones, Sonia Salinas, Eric J Kremer, and Ashley E Pickering. Retrograde optogenetic characterization of the pontospinal module of the locus coeruleus with a canine adenoviral vector. *Brain Research*, 1641(Pt B):274–290, 2016.
- [229] Liang Liang, Alex Fratzl, Glenn Goldey, Rohan N. Ramesh, Arthur U. Sugden, Josh L. Morgan, Chinfei Chen, and Mark L. Andermann. A fine-scale functional logic to convergence from retina to thalamus. *Cell*, 173(6):1343–1355.e24, May 2018.
- [230] Liang Liang, Alex Fratzl, Jasmine D.S. Reggiani, Omar El Mansour, Chinfei Chen, and Mark L. Andermann. Retinal inputs to the thalamus are se-

- lectively gated by arousal. *Current Biology*, 30(20):3923–3934.e9, October 2020.
- [231] Chia-Wei Lin, Shuyin Sim, Alice Ainsworth, Masayoshi Okada, Wolfgang Kelsch, and Carlos Lois. Genetically increased cell-intrinsic excitability enhances neuronal integration into adult brain circuits. *Neuron*, 65(1):32–39, 2010.
- [232] Jack Lindsey, Samuel A. Ocko, Surya Ganguli, and Stephane Deny. A unified theory of early visual representations from retina to cortex through anatomically constrained deep cnns, 2019.
- [233] D Lodge, NA Anis, and NR Burton. Effects of optical isomers of ketamine on excitation of cat and rat spinal neurones by amino acids and acetylcholine. *Neuroscience letters*, 29(3):281–286, 1982.
- [234] Nikos K Logothetis. The underpinnings of the bold functional magnetic resonance imaging signal. *Journal of Neuroscience*, 23(10):3963–3971, 2003.
- [235] Sharon E Loughlin, Stephen L Foote, and Reinhard Grzanna. Efferent projections of nucleus locus coeruleus: Morphologic subpopulations have different efferent targets. *Neuroscience*, 18(2):307–319, 1986.
- [236] Konstantin-Klemens Lurz, Mohammad Bashiri, Konstantin Willeke, Akshay K Jagadish, Eric Wang, Edgar Y Walker, Santiago A Cadena, Taliah Muhammad, Erick Cobos, Andreas S Tolias, et al. Generalization in data-driven models of primary visual cortex. *bioRxiv*, pages 2020–10, 2020.
- [237] Alexandra T Luskin, Liya Li, Xiaoyan Fu, Katherine Barcomb, Trevor Blackburn, Emma M Li, Azka Rana, Rachel C Simon, Lijia Sun, Anna D Murry, Sam A Golden, Garret D Stuber, Christopher P Ford, Lianqing Gu,

- and Michael R Bruchas. A diverse network of pericoerulear neurons control arousal states. *BioRxiv*, 2022(06):30.498327, 2022.
- [238] Linda Madisen, Theresa A Zwingman, Susan M Sunkin, Seung Wook Oh, A Hatim, Hong Gu, Lydia L Ng, Richard D Palmiter, Michael J Hawrylycz, R Allan, Ed S Lein, and Hongkui Zeng. A robust and high-throughput cre reporting and characterization system for the whole mouse brain. *Nature Neuroscience*, 13(1):133–140, 2010.
- [239] Niru Maheswaranathan, Lane T McIntosh, Hidenori Tanaka, Satchel Grant, David B Kastner, Joshua B Melander, Aran Nayebi, Luke E Brezovec, Julia H Wang, Surya Ganguli, et al. Interpreting the retinal neural code for natural scenes: From computations to neurons. *Neuron*, 111(17):2742–2755, 2023.
- [240] Mathias Mahn, Matthias Prigge, Shiri Ron, Rivka Levy, and Ofer Yizhar. Biophysical constraints of optogenetic inhibition at presynaptic terminals. *Nature Neuroscience*, 19:554–556, March 2016. ISSN 1546-1726.
- [241] Mathias Mahn, Lihi Gibor, Pritish Patil, Katayun Cohen-Kashi Malina, Shir Oring, Yoav Printz, Rivka Levy, Ilan Lampl, and Ofer Yizhar. High-efficiency optogenetic silencing with soma-targeted anion-conducting channelrhodopsins. *Nature Communications*, 9(1):4125, 2018.
- [242] Paul R Manger and Oxana Eschenko. The mammalian locus coeruleus complex—consistencies and variances in nuclear organization. *Brain Sciences*, 11(11):1486, 2021.
- [243] Arun Manohar, Andrea L Curtis, Stephen A Zderic, and Rita J Valentino. Brainstem network dynamics underlying the encoding of bladder information. *eLife*, 6:e29917, 2017.

- [244] Valerio Mante, Vincent Bonin, and Matteo Carandini. Functional Mechanisms Shaping Lateral Geniculate Responses to Artificial and Natural Stimuli. *Neuron*, 58(4):625–638, 2008.
- [245] Adam H Marblestone, Greg Wayne, and Konrad P Kording. Toward an integration of deep learning and neuroscience. *Frontiers in computational neuroscience*, 10:94, 2016.
- [246] Eve Marder and Adam L Taylor. Multiple models to capture the variability in biological neurons and networks. *Nature neuroscience*, 14(2):133–138, 2011.
- [247] Stephen Maren, K Luan Phan, and Israel Liberzon. The contextual brain: implications for fear conditioning, extinction and psychopathology. *Nature reviews neuroscience*, 14(6):417–428, 2013.
- [248] Vasilis Z Marmarelis. *Nonlinear dynamic modeling of physiological systems*, volume 10. John Wiley & Sons, 2004.
- [249] M Maroteaux and M Mameli. Cocaine evokes projection-specific synaptic plasticity of lateral habenula neurons. *J Neurosci*, 32:12641–12646, 2012.
- [250] David Marr. *Vision: A computational investigation into the human representation and processing of visual information*. MIT press, 2010.
- [251] Ilkka K. Martikainen, Nora Hagelberg, Satu K. Jääskeläinen, Jarmo Hietala, and Antti Pertovaara. Dopaminergic and serotonergic mechanisms in the modulation of pain: In vivo studies in human brain. *European Journal of Pharmacology*, 834:337–345, 2018. ISSN 0014-2999.
- [252] Ana R O Martins and Robert C Froemke. Coordinated forms of noradrenergic plasticity in the locus coeruleus and primary auditory cortex. *Nature Neuroscience*, 18(10):1483–1492, 2015.

- [253] JS Marvin, BG Borghuis, L Tian, J Cichon, MT Harnett, J Akerboom, et al. An optimized fluorescent probe for visualizing glutamate neurotransmission. *Nat Methods*, 10:162–170, 2013.
- [254] JS Marvin, Y Shimoda, V Magloire, M Leite, T Kawashima, TP Jensen, et al. A genetically encoded fluorescent sensor for in vivo imaging of gaba. *Nat Methods*, 16:763–770, 2019.
- [255] Masayuki Matsumoto and Okihide Hikosaka. Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature*, 447(7148):1111–1115, 2007.
- [256] Masayuki Matsumoto and Okihide Hikosaka. Representation of negative motivational value in the primate lateral habenula. *Nat Neurosci*, 12:77–84, 2009.
- [257] Kerry McAlonan, James Cavanaugh, and Robert H Wurtz. Attentional modulation of thalamic reticular neurons. *Journal of Neuroscience*, 26(16):4444–4450, 2006.
- [258] David McAlpine. Creating a sense of auditory space. *The Journal of physiology*, 566(1):21–28, 2005.
- [259] Jim McBurney-Lin, Greta Vargova, Machhindra Garad, Edward Zagha, and Hongdian Yang. The locus coeruleus mediates behavioral flexibility. *Cell reports*, 41(4), 2022.
- [260] J W McClurkin and R T Marrocco. Visual cortical input alters spatial tuning in monkey lateral geniculate nucleus cells. *The Journal of Physiology*, 348(1):135–152, 1984.
- [261] David A. McCormick. Neurotransmitter actions in the thalamus and cere-

- bral cortex and their role in neuromodulation of thalamocortical activity. *Progress in Neurobiology*, 39(4):337–388, October 1992.
- [262] David A McCormick. Neurotransmitter actions in the thalamus and cerebral cortex and their role in neuromodulation of thalamocortical activity. *Progress in neurobiology*, 39(4):337–388, 1992.
- [263] Autumn McKinney, Meng Hu, Alicia Hoskins, Alireza Mohammadyar, Naveed Naeem, Jialiang Jing, Sanjay S Patel, Bharat R Sheth, and Xiangmin Jiang. Cellular and circuit organization of the locus coeruleus of adult mice. *BioRxiv*, 2022(03):01.481656, 2022.
- [264] Autumn McKinney, Meng Hu, Alicia Hoskins, Alireza Mohammadyar, Naveed Naeem, Jialiang Jing, Sanjay S Patel, Bharat R Sheth, and Xiangmin Jiang. Cellular composition and circuit organization of the locus coeruleus of adult mice. *eLife*, 12:e80100, 2023.
- [265] Mélanie Megemont, Jamie McBurney-Lin, and Han-Ching Yang. Pupil diameter is not an accurate real-time readout of locus coeruleus activity. *eLife*, 11:e70510, 2022.
- [266] Leo Michel, Patricia Molina, and Manuel Mamelì. The behavioral relevance of a modular organization in the lateral habenula. *Neuron*, 2024.
- [267] Patrick Mineault. Typology of neuroai, 2023. URL [https://figshare.com/articles/figure/Typology\\_of\\_neuroAI/21818403](https://figshare.com/articles/figure/Typology_of_neuroAI/21818403).
- [268] Patrick J. Mineault, Elaine Tring, Joshua T. Trachtenberg, and Dario L. Ringach. Enhanced spatial resolution during locomotion and heightened attention in mouse primary visual cortex. *Journal of Neuroscience*, 36(24):6382–6392, June 2016.

- [269] Kenji Mizuseki, Kamran Diba, Eva Pastalkova, and Gyorgy Buzsaki. Hippocampal ca1 pyramidal cells form functionally distinct sublayers. *Nature Neuroscience*, 14(9):1174–1181, 2011.
- [270] A. C. S. Mok and G. J. Mogenson. Effect of electrical stimulation of the septum and the lateral preoptic area on unit activity of the lateral habenular nucleus in the rat. *Brain Research*, 43(2):361–372, 1972.
- [271] Benedek Molnár, Péter Sere, Sándor Bordé, Krisztián Koós, Nikolett Zsigri, Péter Horváth, and Magor L Lőrincz. Cell type-specific arousal-dependent modulation of thalamic activity in the lateral geniculate nucleus. *Cerebral Cortex Communications*, 2(2):tgab020, 2021.
- [272] Benedek Molnár, Péter Sere, Sándor Bordé, Krisztián Koós, Nikolett Zsigri, Péter Horváth, and Magor L Lőrincz. Cell Type-Specific Arousal-Dependent Modulation of Thalamic Activity in the Lateral Geniculate Nucleus. *Cerebral Cortex Communications*, 2(2):tgab020, 03 2021. ISSN 2632-7376.
- [273] Sarah Mondoloni, Manuel Mameli, and Mauro Congiu. Reward and aversion encoding in the lateral habenula for innate and learned behaviours. *Translational psychiatry*, 12(1):3, 2022.
- [274] JS Morris, KA Smith, PJ Cowen, KJ Friston, and Raymond J Dolan. Covariation of activity in habenula and dorsal raphe nuclei following tryptophan depletion. *Neuroimage*, 10(2):163–172, 1999.
- [275] Eslam Mounier, Bassem Abdullah, Hani Mahdi, and Seif Eldawlatly. A deep convolutional visual encoding model of neuronal responses in the lgn. *Brain Informatics*, 8, June 2021.
- [276] James Munkres. Algorithms for the assignment and transportation prob-

- lems. *Journal of the society for industrial and applied mathematics*, 5(1): 32–38, 1957.
- [277] P. C. Murphy and A. M. Sillito. Corticofugal feedback influences the generation of length tuning in the visual pathway. *Nature*, 329(6141):727–729, oct 1987.
- [278] Simon Musall, Matthew T. Kaufman, Ashley L. Juavinett, Steven Gluf, and Anne K. Churchland. Single-trial neural dynamics are dominated by richly varied movements. *Nature Neuroscience*, 22(10):1677–1686, September 2019.
- [279] Miho Nakajima and Michael M Halassa. Thalamic control of functional cortical connectivity. *Current opinion in neurobiology*, 44:127–131, 2017.
- [280] John Ashworth Nelder and Robert WM Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 135(3):370–384, 1972.
- [281] DB Nestvogel and DA McCormick. Visual thalamocortical mechanisms of waking state dependent activity and alpha oscillations. *bioRxiv*, 1–40, 2021.
- [282] Dennis B. Nestvogel and David A. McCormick. Visual thalamocortical mechanisms of waking state-dependent activity and alpha oscillations. *Neuron*, 110(1):120–138.e4, January 2022.
- [283] Joao P Neto, Goncalo Lopes, João Frazão, Joana Nogueira, Pedro Lacerda, Pedro Baião, Annelies Aarts, Ana Andrei, Sami Musa, Elvira Fortunato, Pedro Barquinha, and Adam R Kampff. Validating silicon polytrodes with paired juxtacellular recordings: Method and dataset. *Journal of Neurophysiology*, 116(2):892–903, 2016.
- [284] Ricardo M Neves, Steven van Keulen, Ming Yang, Nikos K Logothetis,

- and Oxana Eschenko. Locus coeruleus phasic discharge is essential for stimulus-induced gamma oscillations in the prefrontal cortex. *Journal of Neurophysiology*, 119(3):904–920, 2018.
- [285] Joanne Ng, Apostolos Papandreou, Simon J Heales, and Manju A Kurian. Monoamine neurotransmitter disorders—clinical advances and future perspectives. *Nature Reviews Neurology*, 11(10):567–584, 2015.
- [286] Cristopher M. Niell and Michael P. Stryker. Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. *Neuron*, 65(4):472–479, 2010.
- [287] Cristopher M Niell and Michael P Stryker. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*, 65(4):472–479, 2010.
- [288] Alvaro Nuno-Perez, Anna Tchenio, Manuel Mameli, and Salvatore Lecca. Lateral habenula gone awry in depression: bridging cellular adaptations with therapeutics. *Frontiers in neuroscience*, 12:485, 2018.
- [289] Amelie Nuno-Perez, Massimo Trusel, A L Lalive, Massimo Congiu, D Gastaldo, A Tchenio, et al. Stress undermines reward-guided cognitive performance through synaptic depression in the lateral habenula. *Neuron*, 109:947–956.e5, 2021.
- [290] Marcel Oberlaender, Christiaan P J de Kock, Reimer M Bruno, Axel Ramirez, Henner S Meyer, Vincent J Dercksen, Moritz Helmstaedter, and Bert Sakmann. Cell type-specific three-dimensional structure of thalamocortical circuits in a column of rat vibrissa cortex. *Cerebral Cortex*, 22(10):2375–2391, 2012.
- [291] Shawn R. Olsen, Dante S. Bortone, Hillel Adesnik, and Massimo

- Scanziani. Gain control by layer six in cortical circuits of vision. *Nature*, 483(7387):47–52, 2012.
- [292] Shawn R Olsen, Dante S Bortone, Hillel Adesnik, and Massimo Scanziani. Gain control by layer six in cortical circuits of vision. *Nature*, 483(7387):47–52, 2012.
- [293] Bruno A Olshausen and David J Field. How close are we to understanding V1? *Neural computation*, 17(8):1665–1699, 2005.
- [294] Cheryl A Osterhout, Christopher R Sterling, David M Chikaraishi, and A William Tank. Induction of tyrosine hydroxylase in the locus coeruleus of transgenic mice in response to stress or nicotine treatment: Lack of activation of tyrosine hydroxylase promoter activity. *Journal of Neurochemistry*, 94(3):731–741, 2005.
- [295] Marius Pachitariu, Nicholas Steinmetz, Shabnam Kadir, Matteo Carandini, and Kenneth D Harris. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv*, page 061481, 2016.
- [296] J Palacios-Filardo and JR Mellor. Neuromodulation of hippocampal long-term synaptic plasticity. *Curr Opin Neurobiol*, 54:37–43, 2019.
- [297] Liam Paninski. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15(4):243–262, January 2004.
- [298] Liam Paninski, Jonathan Pillow, and Jeremy Lewi. *Statistical models for neural encoding, decoding, and optimal stimulus design*, pages 493–507. Elsevier, 2007. ISBN 9780444528230.
- [299] Stefano Panzeri, Christopher D Harvey, Eugenio Piasini, Peter E Latham, and Tommaso Fellin. Cracking the neural code for sensory perception by

- combining statistics, intervention, and behavior. *Neuron*, 93(3):491–507, 2017.
- [300] Il Memming Park, Miriam L R Meister, Alexander C Huk, and Jonathan W Pillow. Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nature Neuroscience*, 17(10):1395–1403, August 2014.
- [301] Philip R.L. Parker, Elliott T.T. Abe, Emmalyn S.P. Leonard, Dylan M. Martins, and Cristopher M. Niell. Joint coding of visual input and eye/head position in v1 of freely moving mice. *Neuron*, 110(23):3897–3906.e5, December 2022.
- [302] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572, 1901.
- [303] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. URL <http://jmlr.org/papers/v12/pedregosa11a.html>.
- [304] Markku Penttonen, Anita Kamondi, Attila Sik, László Acsády, and Gyorgy Buzsaki. Feed-forward and feed-back activation of the dentate gyrus in vivo during dentate spikes and sharp wave bursts. *Hippocampus*, 7(4):437–450, 1997.
- [305] MA Penzo, V Robert, J Tucciarone, D De Bundel, M Wang, L Van Aelst,

- et al. The paraventricular thalamus controls a central amygdala fear circuit. *Nature*, 519:455–459, 2015.
- [306] Virginia M Pickel, Tong H Joh, and Donald J Reis. Ultrastructural localization of tyrosine hydroxylase in noradrenergic neurons of brain. *Proceedings of the National Academy of Sciences*, 72(2):659–663, 1975.
- [307] Anirban G Pillai, Douwe de Jong, Stelios Kanatsou, Harm Krugers, Anne Knapman, Jean-Marc Heinzmann, Florian Holsboer, Rainer Landgraf, Marian Joëls, and Carsten Touma. Dendritic morphology of hippocampal and amygdalar neurons in adolescent mice is resilient to genetic differences in stress reactivity. *PLoS ONE*, 7(6):e38971, 2012.
- [308] Jonathan W. Pillow, Liam Paninski, Valerie J. Uzzell, Eero P. Simoncelli, and E. J. Chichilnisky. Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *Journal of Neuroscience*, 25(47):11003–11013, 2005.
- [309] Jonathan W. Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M. Litke, E. J. Chichilnisky, and Eero P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, aug 2008.
- [310] Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, EJ Chichilnisky, and Eero P Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.
- [311] Didier Pinault. A novel single-cell staining procedure performed in vivo under electrophysiological control: morpho-functional features of juxta-

- cellularly labeled thalamic cells and other central neurons with biocytin or Neurobiotin. *Journal of Neuroscience Methods*, 65(2):113–136, 1996.
- [312] Didier Pinault. A novel single-cell staining procedure performed in vivo under electrophysiological control: Morpho-functional features of juxta-cellularly labeled thalamic cells and other central neurons with biocytin or neurobiotin. *Journal of Neuroscience Methods*, 65(2):113–136, 1996.
- [313] D. M. Piscopo, R. N. El-Danaf, A. D. Huberman, and C. M. Niell. Diverse Visual Features Encoded in Mouse Lateral Geniculate Nucleus. *Journal of Neuroscience*, 33(11):4642–4656, 2013.
- [314] Nicholas W Plummer, Elisabetta L Scappini, Kevin G Smith, Chad J Tucker, and Per Jensen. Two subpopulations of noradrenergic neurons in the locus coeruleus complex distinguished by expression of the dorsal neural tube marker pax7. *Frontiers in Neuroanatomy*, 11:60, 2017.
- [315] Gina R Poe, S Foote, Oxana Eschenko, Joshua P Johansen, Sebastien Bouret, Gary Aston-Jones, Christopher W Harley, Denise Manahan-Vaughan, David Weinshenker, Rita Valentino, Craig Berridge, Daniel J Chandler, Barry Waterhouse, and Susan J Sara. Locus coeruleus: A new look at the blue spot. *Nature Reviews Neuroscience*, 21(11):644–659, 2020.
- [316] RB Poorthuis, L Enke, and JJ Letzkus. Cholinergic circuit modulation through differential recruitment of neocortical interneuron types during behaviour. *J Physiol*, 592:4155–4164, 2014.
- [317] Alexandre Pouget, Peter Dayan, and Richard Zemel. Information processing with population codes. *Nature Reviews Neuroscience*, 1(2):125–132, 2000.
- [318] James F A Poulet, Laura M J Fernandez, Sylvain Crochet, and Carl C H

- Petersen. Thalamic control of cortical states. *Nature Neuroscience*, 15(3): 370–372, January 2012.
- [319] C. A. Poynton. Rehabilitation of gamma. In B. E. Rogowitz and T. N. Pappas, editors, *Human Vision and Electronic Imaging III*, volume 3299, pages 232–249, San Jose, CA, 1998. International Society for Optical Engineering.
- [320] Pablo Preston-Ferrer, Silvia Coletta, Martin Frey, and Andrea Burgalossi. Anatomical organization of presubicular head-direction circuits. *eLife*, 5: e14592, 2016.
- [321] Christophe D Proulx, Okihide Hikosaka, and Roberto Malinow. Reward processing by the lateral habenula in normal and depressive behaviors. *Nature neuroscience*, 17(9):1146–1152, 2014.
- [322] Rodrigo Quian Quiroga. Spike sorting. *Current Biology*, 22(2):R45–R46, 2012.
- [323] Jacob Reimer, Emmanouil Froudarakis, Cathryn R. Cadwell, Dimitri Yatsenko, George H. Denfield, and Andreas S. Tolias. Pupil Fluctuations Track Fast Switching of Cortical States during Quiet Wakefulness. *Neuron*, 84(2): 355–362, oct 2014.
- [324] Kimberly Reinhold, Arbora Resulaj, and Massimo Scanziani. Brain state-dependent modulation of thalamic visual processing by cortico-thalamic feedback. *The Journal of Neuroscience*, 43(9):1540–1554, January 2023.
- [325] Pieter R Roelfsema and Floris P de Lange. Early visual cortex as a multiscale cognitive blackboard. *Annual review of vision science*, 2:131–151, 2016.
- [326] Miroslav Román Rosón, Yannik Bauer, Ann Kotkat, Philipp Berens,

- Thomas Euler, and Laura Busse. Mouse dLGN Receives Functional Input from a Diverse Population of Retinal Ganglion Cells with Limited Convergence. *Neuron*, 102(2):462–476, 2019.
- [327] MA Rossi, ML Basiri, Y Liu, Y Hashikawa, K Hashikawa, LE Fenno, et al. Transcriptional and functional divergence in lateral hypothalamic glutamate neurons projecting to the lateral habenula and ventral tegmental area. *Neuron*, 109:3823–3837.e6, 2021.
- [328] Lisa Roux, Eran Stark, Lucas Sjulson, and Gyorgy Buzsaki. In vivo optogenetic identification and manipulation of gabaergic interneuron subtypes. *Current Opinion in Neurobiology*, 26:88–95, 2014.
- [329] Sebastien Royer, Boris V Zemelman, Mladen Barbic, Attila Losonczy, Gyorgy Buzsaki, and Jeffrey C Magee. Multi-array silicon probes with integrated optical fibers: Light-assisted perturbation and recording of local neural circuits in the behaving animal. *European Journal of Neuroscience*, 31(12):2279–2291, 2010.
- [330] Curtis T. Rueden, Johannes Schindelin, Mark C. Hiner, Barry E. DeZonia, Alison E. Walter, Ellen T. Arena, and Kevin W. Eliceiri. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinf.*, 18(1), December 2017. ISSN 1471-2105.
- [331] Caroline A. Runyan, Eugenio Piasini, Stefano Panzeri, and Christopher D. Harvey. Distinct timescales of population coding across cortex. *Nature*, 548(7665):92–96, July 2017.
- [332] Nicole C Rust, Odelia Schwartz, J Anthony Movshon, and Eero P Simoncelli. Spatiotemporal elements of macaque v1 receptive fields. *Neuron*, 46(6):945–956, 2005.

- [333] Yuri B Saalman and Sabine Kastner. Cognitive and perceptual functions of the visual thalamus. *Neuron*, 71(2):209–223, 2011.
- [334] Helia Safaai, Ricardo Neves, Oxana Eschenko, Nikos K Logothetis, and Stefano Panzeri. Modeling the effect of locus coeruleus firing on cortical state dynamics and single-trial sensory processing. *Proceedings of the National Academy of Sciences*, 112(41):12834–12839, 2015.
- [335] Ramiro Salas, Philip Baldwin, Mariella De Biasi, and Read Montague. Bold responses to negative reward prediction errors in human habenula. *Frontiers in human neuroscience*, 4:1596, 2010.
- [336] Ebony R Samuels and Elemer Szabadi. Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part i: principles of functional organisation. *Current neuropharmacology*, 6(3):235–253, 2008.
- [337] Alessandro Sanzeni, Agostina Palmigiano, Tuan H. Nguyen, Junxiang Luo, Jonathan J. Nassi, John H. Reynolds, Mark H. Histed, Kenneth D. Miller, and Nicolas Brunel. Mechanisms underlying reshuffling of visual responses by optogenetic stimulation in mice and monkeys. *Neuron*, 111(24):4102–4115.e9, December 2023.
- [338] Susan J Sara. The locus coeruleus and noradrenergic modulation of cognition. *Nature reviews neuroscience*, 10(3):211–223, 2009.
- [339] Susan J Sara. Locus coeruleus in time with the making of memories. *Current Opinion in Neurobiology*, 35:87–94, 2015.
- [340] Susan J Sara and Sebastien Bouret. Orienting and reorienting: The locus coeruleus mediates cognition through arousal. *Neuron*, 76(1):130–141, 2012.

- [341] Susan J Sara and Menahem Segal. Plasticity of sensory responses of locus coeruleus neurons in the behaving rat: Implications for cognition. In *Progress in Brain Research*, volume 88, pages 571–585. Elsevier, 1991.
- [342] Alexander Sartorius, Karl L Kiening, Peter Kirsch, Carl C von Gall, Uwe Haberkorn, Andreas W Unterberg, Fritz A Henn, and Andreas Meyer-Lindenberg. Remission of major depression under deep brain stimulation of the lateral habenula in a therapy-refractory patient. *Biological psychiatry*, 67(2):e9–e11, 2010.
- [343] Arpiar Saunders, Evan Z Macosko, Alec Wysoker, Melissa Goldman, Fenna M Krienen, Heather de Rivera, Elizabeth Bien, Matthew Baum, Laura Bortolin, Shuyu Wang, et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell*, 174(4):1015–1030, 2018.
- [344] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, Jean-Yves Tinevez, Daniel J White, Volker Hartenstein, Kevin Eliceiri, Pavel Tomancak, and Albert Cardona. Fiji: An open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682, 2012.
- [345] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, Jean-Yves Tinevez, Daniel James White, Volker Hartenstein, Kevin Eliceiri, Pavel Tomancak, and Albert Cardona. Fiji: An open-source platform for biological-image analysis. *Nat. Methods*, 9(7):676–682, July 2012. ISSN 1548-7091, 1548-7105.
- [346] Kathleen Schmidt, Bahareh Bari, Martina Ralle, Cheryl Washington-

- Hughes, Achchhe Muchenditsi, Elizabeth Maxey, and Svetlana Lutsenko. Localization of the locus coeruleus in the mouse brain. *Journal of Visualized Experiments*, 145:e58652, 2019.
- [347] Wolfram Schultz, Peter Dayan, and P Read Montague. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599, 1997.
- [348] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.
- [349] Loren A Schwarz, Kazunari Miyamichi, Xiaojing J Gao, Kevin T Beier, Ben Weissbourd, Katherine E DeLoach, Junchol Ren, Sandra Ibanes, Robert C Malenka, Eric J Kremer, and Liqun Luo. Viral-genetic tracing of the input-output organization of a central norepinephrine circuit. *Nature*, 524(7563):88–92, 2015.
- [350] Caspar M Schwiedrzik, Wolf Singer, and Lucia Melloni. Subjective and objective learning effects dissociate in space and in time. *Proceedings of the National Academy of Sciences*, 108(11):4506–4511, 2011.
- [351] Yuta Senzai and György Buzsáki. Physiological properties and behavioral correlates of hippocampal granule cells and mossy cells. *Neuron*, 93(3):691–704.e5, 2017.
- [352] GH Seol, J Ziburkus, S Huang, L Song, IT Kim, K Takamiya, et al. Neuromodulators control the polarity of spike-timing-dependent synaptic plasticity. *Neuron*, 55:919–929, 2007.
- [353] H Sebastian Seung and Uygur Sümbül. Neuronal cell types and connectivity: lessons from the retina. *Neuron*, 83(6):1262–1272, 2014.
- [354] SJ Shabel, CD Proulx, A Trias, RT Murphy, and R Malinow. Input to

- the lateral habenula from the basal ganglia is excitatory, aversive, and suppressed by serotonin. *Neuron*, 74:475–481, 2012.
- [355] SJ Shabel, CD Proulx, J Piriz, and R Malinow. Mood regulation. gaba/glutamate co-release controls habenula output and is modified by antidepressant treatment. *Science*, 345:1494–1498, 2014.
- [356] Michael N Shadlen and William T Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of neuroscience*, 18(10):3870–3896, 1998.
- [357] Shawn D Shea, Lawrence C Katz, and Richard Mooney. Noradrenergic induction of odor-specific neural habituation and olfactory memories. *Journal of Neuroscience*, 28(42):10711–10719, 2008.
- [358] L Shelton, L Becerra, and D Borsook. Unmasking the mysteries of the habenula in pain and analgesia. *Progress in neurobiology*, 96(2):208–219, 2012.
- [359] L Shelton, G Pendse, N Maleki, EA Moulton, A Lebel, L Becerra, et al. Mapping pain activation and connectivity of the human habenula. *J Neurophysiol*, 107:2633–2648, 2012.
- [360] Gordon M. G. Shepherd and Naoki Yamawaki. Untangling the cortico-thalamo-cortical loop: cellular pieces of a knotty circuit puzzle. *Nature Reviews Neuroscience*, 22(7):389–406, May 2021.
- [361] S. Murray Sherman. Tonic and burst firing: Dual modes of thalamocortical relay. *Trends in Neurosciences*, 24(2):122–126, 2001.
- [362] S Murray Sherman and R W Guillery. The role of the thalamus in the flow of information to the cortex. *Philosophical transactions of the Royal Society of London*, 357(1428):1695–708, 2002.

- [363] S Murray Sherman and RW Guillery. On the actions that one nerve cell can have on another: distinguishing “drivers” from “modulators”. *Proceedings of the National Academy of Sciences*, 95(12):7121–7126, 1998.
- [364] SM Sherman and Christof Koch. The control of retinogeniculate transmission in the mammalian lateral geniculate nucleus. *Experimental Brain Research*, 63:1–20, 1986.
- [365] Michael T Shipley, Li Fu, Matthew Ennis, Wei Liu, and Gary Aston-Jones. Dendrites of locus coeruleus neurons extend preferentially into two pericoerulear zones. *Journal of Comparative Neurology*, 365(1):56–68, 1996.
- [366] Joshua H Siegle, Xiaoxuan Jia, Séverine Durand, Sam Gale, Corbett Bennett, Nile Graddis, Gregory Heller, Tamina K Ramirez, Hannah Choi, Jennifer A Luviano, et al. Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, 592(7852):86–92, 2021.
- [367] Adam M. Sillito and Helen E. Jones. Corticothalamic interactions in the transfer of visual information. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1428):1739–1752, dec 2002.
- [368] Adam M. Sillito, Javier Cudeiro, and Helen E. Jones. Always returning: feedback and sensory processing in visual cortex and thalamus. *Trends in Neurosciences*, 29(6):307–316, 2006.
- [369] Adam M Sillito, Javier Cudeiro, and Helen E Jones. Always returning: feedback and sensory processing in visual cortex and thalamus. *Trends in neurosciences*, 29(6):307–316, 2006.
- [370] Eero P Simoncelli and Bruno A Olshausen. Natural image statistics and

- neural representation. *Annual review of neuroscience*, 24(1):1193–1216, 2001.
- [371] Amit Singhal et al. Modern information retrieval: A brief overview. *IEEE Data Eng. Bull.*, 24(4):35–43, 2001.
- [372] Anton Sirota and György Buzsáki. Interaction between neocortical and hippocampal networks via slow oscillations. *Thalamus & Related Systems*, 3(4):245–259, 2005.
- [373] William R Softky and Christof Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *Journal of neuroscience*, 13(1):334–350, 1993.
- [374] AT Sørensen, YA Cooper, MV Baratta, FJ Weng, Y Zhang, K Ramamoorthi, et al. A robust activity marking system for exploring active neuronal ensembles. *Elife*, 5:e13918, 2016.
- [375] Francisco Sotres-Bayon and Gregory J Quirk. Prefrontal control of fear: more than just extinction. *Current opinion in neurobiology*, 20(2):231–235, 2010.
- [376] M. A. Spacek, T. J. Blanche, and N. V. Swindale. Python for large-scale electrophysiology. *Front. Neuroinform.*, 2:9, 2009.
- [377] Martin A Spacek, Davide Crombie, Yannik Bauer, Gregory Born, Xinyu Liu, Steffen Katzner, and Laura Busse. Robust effects of corticothalamic feedback and behavioral state on movie responses in mouse dLGN. *Elife*, 11:e70469, 2022.
- [378] Martin A Spacek, Davide Crombie, Yannik Bauer, Gregory Born, Xinyu Liu, Steffen Katzner, and Laura Busse. Robust effects of corticothalamic

- feedback and behavioral state on movie responses in mouse dlgn. *Elife*, 11: e70469, 2022.
- [379] Nicholas C Spitzer. Neurotransmitter switching? no surprise. *Neuron*, 86(5):1131–1144, 2015.
- [380] Eran Stark, Tibor Koos, and György Buzsáki. Diode probes for spatiotemporal optical control of multiple neurons in freely moving animals. *Journal of Neurophysiology*, 108(1):349–363, 2012.
- [381] Nicholas A. Steinmetz, Peter Zatka-Haas, Matteo Carandini, and Kenneth D. Harris. Distributed coding of choice, action and engagement across the mouse brain. *Nature*, 576(7786):266–273, November 2019.
- [382] Marcus Stephenson-Jones, Kevin Yu, Sandra Ahrens, Jason M Tucciarone, A. N van Huijstee, Luis A Mejia, et al. A basal ganglia circuit for evaluating action outcomes. *Nature*, 539:289–293, 2016.
- [383] WC Stern, A Johnson, JD Bronzino, and PJ Morgane. Effects of electrical stimulation of the lateral habenula on single-unit activity of raphe neurons. *Experimental neurology*, 65(2):326–342, 1979.
- [384] Charles F Stevens. Neuronal diversity: too many cell types for comfort? *Current Biology*, 8(20):R708–R710, 1998.
- [385] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Charu Bai Reddy, Matteo Carandini, and Kenneth D. Harris. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 364(6437):eaav7893, 2019.
- [386] Zeng Su and Joshua Y Cohen. Two types of locus coeruleus norepinephrine neurons drive reinforcement learning. *BioRxiv*, 2022(12):08.519670, 2022.
- [387] Daisuke Sugiyama, Sun Woo Hur, Ashley E Pickering, Daichi Kase, Se-

- ung Ju Kim, Mitsuhiro Kawamata, Keiji Imoto, and Hidemasa Furue. In vivo patch-clamp recording from locus coeruleus neurones in the rat brainstem. *Journal of Physiology*, 590:2225–2231, 2012.
- [388] Robert J Sutherland. The dorsal diencephalic conduction system: a review of the anatomy and functions of the habenular complex. *Neuroscience & Biobehavioral Reviews*, 6(1):1–13, 1982.
- [389] H. A. Swadlow and T. G. Weyand. Corticogeniculate neurons, corticotectal neurons, and suspected interneurons in visual cortex of awake rabbits: receptive-field properties, axonal properties, and effects of EEG arousal. *Journal of Neurophysiology*, 57(4):977–1001, April 1987.
- [390] Larry W Swanson. The locus coeruleus: A cytoarchitectonic, golgi and immunohistochemical study in the albino rat. *Brain Research*, 110(1):39–56, 1976.
- [391] Larry W Swanson and Boyd K Hartman. The central adrenergic system. an immunofluorescence study of the location of cell bodies and their efferent connections in the rat utilizing dopamine-b-hydroxylase as a marker. *Journal of Comparative Neurology*, 163:467–505, 1975.
- [392] Kaitlin M Swift, Beth A Gross, Michelle A Frazer, Devin S Bauer, Kevin JD Clark, Elena M Vazey, Gary Aston-Jones, Yulan Li, Ashley E Pickering, Susan J Sara, and Gina R Poe. Abnormal locus coeruleus sleep activity alters sleep signatures of memory consolidation and impairs place cell stability and spatial memory. *Current Biology*, 28(22):3599–3609.e4, 2018.
- [393] Nicholas V. Swindale and Martin A. Spacek. Spike sorting for polytrodes:

- a divide and conquer approach. *Frontiers in Systems Neuroscience*, 8, 2014. ISSN 1662-5137.
- [394] Emily L Sylwestrak, Youngsik Jo, Sam Vesuna, Xiaoyin Wang, Brian Holcomb, Robert H Tien, et al. Cell-type-specific population dynamics of diverse reward computations. *Cell*, 185:3568–3587.e27, 2022.
- [395] Kenji Takahashi, Yuji Kayama, Jiang-Shan Lin, and Kenji Sakai. Locus coeruleus neuronal activity during the sleep-waking cycle in mice. *Neuroscience*, 169(3):1115–1126, 2010.
- [396] Jiaying Tang, Silvia C. Ardila Jimenez, Subhojit Chakraborty, and Simon R. Schultz. Visual receptive field properties of neurons in the mouse lateral geniculate nucleus. *PLOS ONE*, 11(1):e0146017, jan 2016.
- [397] Etienne Tanguay, Sebastien-J Bouchard, Maxime Lévesque, Pierre De Koninck, and Vincent Breton-Provencher. Shining light on the noradrenergic system. *Neurophotonics*, 10(4):044406, 2023.
- [398] Kaycie K. Tayler, Kazumasa Z. Tanaka, Leon G. Reijmers, and Brian J. Wiltgen. Reactivation of neural ensembles during the retrieval of recent and remote memory. *Current Biology*, 23(2):99–106, 2013.
- [399] Frédéric E Theunissen, Stephen V David, Nandini C Singh, Anne Hsu, William E Vinje, and Jack L Gallant. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network: Computation in Neural Systems*, 12(3):289, 2001.
- [400] Alexander Thiele. Optimizing brain processing. *Nature neuroscience*, 12(11):1359–1360, 2009.
- [401] Ju Tian and Naoshige Uchida. Habenula lesions reveal that multiple mech-

- anisms underlie dopamine prediction errors. *Neuron*, 87(6):1304–1316, 2015.
- [402] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.
- [403] Rylie P Tillage, Natale R Sciolino, Nicholas W Plummer, David Lustberg, Lauren C Liles, Ming Hsiang, Justin M Powell, Kevin G Smith, Per Jensen, and David Weinshenker. Elimination of galanin synthesis in noradrenergic neurons reduces galanin in select brain areas and promotes active coping behaviors. *Brain Structure and Function*, 225(2):785–803, 2020.
- [404] Lindsay S Tortorelli, Jamie McBurney-Lin, Muna Garad, Mélanie Mege-mont, Miriam Riccomagno, and Han-Ching Yang. Physiological and functional heterogeneity in the mouse locus coeruleus. *BioRxiv*, 2022(08): 19.504582, 2022.
- [405] Nelson KB Totah, Ricardo M Neves, Stefano Panzeri, Nikos K Logothetis, and Oxana Eschenko. The locus coeruleus is a complex and differentiated neuromodulatory system. *Neuron*, 99(5):1055–1068.e6, 2018.
- [406] Nelson KB Totah, Nikos K Logothetis, and Oxana Eschenko. Noradren-ergic ensemble-based modulation of cognition over multiple timescales. *Brain Research*, 1709:50–66, 2019.
- [407] Jon Touryan, Brian Lau, and Yang Dan. Isolation of relevant visual features from random stimuli for cortical complex cells. *Journal of Neuroscience*, 22(24):10811–10818, 2002.
- [408] NF Trojanowski, J Bottorff, and GG Turrigiano. Activity labeling in vivo

using campari2 reveals intrinsic and synaptic differences between neurons with high and low firing rate set points. *Neuron*, 109:663–676.e5, 2021.

- [409] Massimo Trusel, Anna Cavaccini, Marta Gritti, Barbara Greco, Pierre-Philippe Saintot, Cristiano Nazzaro, Milica Cerovic, Ilaria Morella, Riccardo Brambilla, and Raffaella Tonini. Coordinated regulation of synaptic plasticity at striatopallidal and striatonigral neurons orchestrates motor control. *Cell reports*, 13(7):1353–1365, 2015.
- [410] Massimo Trusel, A Nuno-Perez, Salvatore Lecca, Hiroshi Harada, A L Lalive, M Congiu, et al. Punishment-predictive cues guide avoidance through potentiation of hypothalamus-to-habenula synapses. *Neuron*, 102:120–127.e4, 2019.
- [411] T. Tsumoto, O.D. Creutzfeldt, and C.R. Legendy. Functional organization of the corticofugal system from visual cortex to lateral geniculate nucleus in the cat: With an appendix on geniculo-cortical mono-synaptic connections. *Experimental Brain Research*, 32(3), July 1978.
- [412] Atsushi Uematsu, Bogdan Z Tan, Edgar A Ycu, Javier S Cuevas, Johanna Koivumaa, Felix Junyent, Eric J Kremer, Ilana B Witten, Karl Deisseroth, and Joshua P Johansen. Modular organization of the brainstem noradrenaline system coordinates opposing learning states. *Nature Neuroscience*, 20(11):1602–1611, 2017.
- [413] Markus Ullsperger and D Yves Von Cramon. Error monitoring using external feedback: specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. *Journal of Neuroscience*, 23(10):4308–4314, 2003.
- [414] W. M. Usrey and H. J. Alitto. Visual Functions of the Thalamus. *Annual*

- Review of Vision Science*, 1(1):351–371, 2015.
- [415] W. Martin Usrey, John B. Reppas, and R. Clay Reid. Paired-spike interactions and synaptic efficacy of retinal inputs to the thalamus. *Nature*, 395(6700):384–387, September 1998.
- [416] Martin Vinck, Renata Batista-Brito, Ulf Knoblich, and Jessica A. Cardin. Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. *Neuron*, 86(3):740–754, may 2015.
- [417] William E Vinje and Jack L Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456):1273–1276, 2000.
- [418] Atsushi Wagatsuma, Takashi Okuyama, Chengzhe Sun, Lindsey M Smith, Ko Abe, and Susumu Tonegawa. Locus coeruleus input to hippocampal ca3 drives single-trial learning of a novel context. *Proceedings of the National Academy of Sciences*, 115(2):E310–E316, 2018.
- [419] Franziska Wagner, Torsten Weiss, and Rüdiger W Veh. Electrophysiological properties of neurons and synapses in the lateral habenular complex (lhb). *Pharmacology Biochemistry and Behavior*, 162:38–45, 2017.
- [420] Edgar Y. Walker, Fabian H. Sinz, Erick Cobos, Taliah Muhammad, Emmanouil Froudarakis, Paul G. Fahey, Alexander S. Ecker, Jacob Reimer, Xaq Pitkow, and Andreas S. Tolias. Inception loops discover what excites neurons most using deep predictive models. *Nature Neuroscience*, 22(12):2060–2065, November 2019.
- [421] Edgar Y Walker, Fabian H Sinz, Erick Cobos, Taliah Muhammad, Emmanouil Froudarakis, Paul G Fahey, Alexander S Ecker, Jacob Reimer, Xaq Pitkow, and Andreas S Tolias. Inception loops discover what excites

- neurons most using deep predictive models. *Nature neuroscience*, 22(12): 2060–2065, 2019.
- [422] Michael L Wallace, Kee Wui Huang, Daniel Hochbaum, Minsuk Hyun, Gianna Radeljic, and Bernardo L Sabatini. Anatomical and single-cell transcriptional profiling of the murine habenular complex. *Elife*, 9:e51271, 2020.
- [423] J Wan, W Peng, X Li, T Qian, K Song, J Zeng, et al. A genetically encoded sensor for measuring serotonin dynamics. *Nat Neurosci*, 24:746–752, 2021.
- [424] Daqing Wang, Yi Li, Qiru Feng, Qingchun Guo, Jingfeng Zhou, and Minmin Luo. Learning shapes the aversion and reward responses of lateral habenula neurons. *Elife*, 6:e23045, 2017.
- [425] Rex Y Wang and George K Aghajanian. Physiological evidence for habenula as major link between forebrain and midbrain raphe. *Science*, 197(4298):89–91, 1977.
- [426] Xin-Yue Wang, Xiang Xu, Rui Chen, Wen-Bin Jia, Peng-Fei Xu, Xiao-Qing Liu, Ying Zhang, Xin-Feng Liu, and Yan Zhang. The thalamic reticular nucleus-lateral habenula circuit regulates depressive-like behaviors in chronic stress and chronic pain. *Cell Reports*, 42(10), 2023.
- [427] David K Warland, Pamela Reinagel, and Markus Meister. Decoding visual information from a population of retinal ganglion cells. *Journal of neurophysiology*, 78(5):2336–2350, 1997.
- [428] Alison I. Weber and Jonathan W. Pillow. Capturing the dynamical repertoire of single neurons with generalized linear models. *Neural Computation*, 29(12):3260–3289, December 2017.
- [429] J. F. Webster, R. Vroman, K. Balueva, et al. Disentangling neuronal inhibi-

- tion and inhibitory pathways in the lateral habenula. *Scientific Reports*, 10: 8490, 2020.
- [430] T. Weiss and R. W. Veh. Morphological and electrophysiological characteristics of neurons within identified subnuclei of the lateral habenula in rat brain slices. *Neuroscience*, 172:74–93, 2011.
- [431] Caroline HK West, James C Ritchie, Katherine A Boss-Williams, and James M Weiss. Antidepressant drugs with differing pharmacological actions decrease activity of locus coeruleus neurons. *International Journal of Neuropsychopharmacology*, 12(5):627–641, 2009.
- [432] Theodore G. Weyand. The multifunctional lateral geniculate nucleus. *Reviews in the Neurosciences*, 27(2):135–157, feb 2016.
- [433] Theodore G Weyand. The multifunctional lateral geniculate nucleus. *Reviews in the Neurosciences*, 27(2):135–157, 2016.
- [434] J Simon Wiegert and Thomas G Oertner. How (not) to silence long-range projections with light. *Nature Neuroscience*, 19(4):527–528, 2016.
- [435] J. Simon Wiegert, Mathias Mahn, Matthias Prigge, Yoav Printz, and Ofer Yizhar. Silencing Neurons: Tools, Applications, and Experimental Constraints. *Neuron*, 95(3):504–529, 2017.
- [436] Jonas Wietek, Silvia Rodriguez-Rozada, Janine Tutas, Federico Tenedini, Christiane Grimm, Thomas G. Oertner, Peter Soba, Peter Hegemann, and J. Simon Wiegert. Anion-conducting channelrhodopsins with tuned spectra and modified kinetics engineered for optogenetic manipulation of behavior. *Scientific Reports*, 7(1):1–18, 2017.
- [437] K. S. Wilcox, M. J. Gutnick, and G. R. Christoph. Electrophysiological

- properties of neurons in the lateral habenula nucleus: an in vitro study. *Journal of Neurophysiology*, 59(1):212–225, 1988.
- [438] John Williams, Graeme Henderson, and Richard North. Locus coeruleus neurons. In *A2-Adrenoceptor Agonists*, pages 297–311. Springer, 1984.
- [439] Robert W Williams and Karl Herrup. The control of neuron number. *Annual review of neuroscience*, 11(1):423–453, 1988.
- [440] Jakob Wolfart, Damien Debay, Gwendal Le Masson, Alain Destexhe, and Thierry Bal. Synaptic background activity controls spike transfer from thalamus to cortex. *Nature Neuroscience*, 8(12):1760–1767, 2005.
- [441] CIC Wolfe, EK Hwang, EC Ijomor, A Zapata, AF Hoffman, and CR Lupica. Muscarinic acetylcholine m2 receptors regulate lateral habenula neuron activity and control cocaine seeking behavior. *J Neurosci*, 42: 5552–5563, 2022.
- [442] Michael C-K Wu, Stephen V David, and Jack L Gallant. Complete functional characterization of sensory neurons by system identification. *Annu. Rev. Neurosci.*, 29:477–505, 2006.
- [443] Keith Wynroe, David Atkinson, and Jaime Sevilla. Literature review of transformative artificial intelligence timelines, 2023. URL <https://epochai.org/blog/>. Accessed: 2024-07-26.
- [444] Lu Xiang, Alon Harel, Haiyan Gao, Ashley E Pickering, Susan J Sara, and Sidney I Wiener. Behavioral correlates of activity of optogenetically identified locus coeruleus noradrenergic neurons in rats performing t-maze tasks. *Scientific Reports*, 9(1):1361, 2019.
- [445] Miyuki Yamasaki and Takashi Takeuchi. Locus coeruleus and dopamine-dependent memory consolidation. *Neural Plasticity*, 2017:8602690, 2017.

- [446] Daniel LK Yamins and James J DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, 19(3):356–365, 2016.
- [447] Yan Yang, Yihui Cui, Kangning Sang, Yiyan Dong, Zheyi Ni, Shuangshuang Ma, and Hailan Hu. Ketamine blocks bursting in the lateral habenula to rapidly relieve depression. *Nature*, 554(7692):317–322, 2018.
- [448] Yan Yang, Hao Wang, Ji Hu, and Hailan Hu. Lateral habenula in the pathophysiology of depression. *Current opinion in neurobiology*, 48:90–96, 2018.
- [449] Lina Yassin, Brett L. Benedetti, Jean-Sébastien Jouhanneau, Jing A. Wen, James F.A. Poulet, and Alison L. Barth. An embedded subnetwork of highly active neurons in the neocortex. *Neuron*, 68(6):1043–1050, 2010.
- [450] Dimitri Yatsenko, Edgar Y. Walker, and Andreas S. Tolias. DataJoint: A simpler relational data model, 2018.
- [451] XD Yu, Y Zhu, QX Sun, F Deng, J Wan, D Zheng, et al. Distinct serotonergic pathways to the amygdala underlie separate behavioral features of anxiety. *Nat Neurosci*, 25:1651–1663, 2022.
- [452] Rafael Yuste. From the neuron doctrine to neural networks. *Nature reviews neuroscience*, 16(8):487–497, 2015.
- [453] Anthony Zador, Sean Escola, Blake Richards, Bence Ölveczky, Yoshua Bengio, Kwabena Boahen, Matthew Botvinick, Dmitri Chklovskii, Anne Churchland, Claudia Clopath, et al. Catalyzing next-generation artificial intelligence through neuroai. *Nature communications*, 14(1):1597, 2023.
- [454] A Zapata, EK Hwang, and CR Lupica. Lateral habenula involvement in impulsive cocaine seeking. *Neuropsychopharmacology*, 42:1103–1112, 2017.

- [455] Hongkui Zeng and Joshua R Sanes. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nature Reviews Neuroscience*, 18(9):530–546, 2017.
- [456] Siyu Zhang, Min Xu, Tsukasa Kamigaki, Johnny Phong Hoang Do, Wei-Cheng Chang, Sean Jenvay, Kazunari Miyamichi, Liquan Luo, and Yang Dan. Long-range and local circuits for top-down modulation of visual cortex processing. *Science*, 345(6197):660–665, August 2014.
- [457] Xiaomin Zhang and Peter Jonas. Integration of spatial and non-spatial information by heterogeneous dentate gyrus granule cells. *Journal of Life Sciences (Westlake Village, Calif.)*, 2(4):19, 2020.
- [458] Hua Zhao, Bei-Lin Zhang, Shao-Jun Yang, and Benjamin Rusak. The role of lateral habenula–dorsal raphe nucleus circuits in higher brain functions and psychiatric illness. *Behavioural brain research*, 277:89–98, 2015.
- [459] Z Zheng, C Guo, M Li, L Yang, P Liu, X Zhang, et al. Hypothalamus–habenula potentiation encodes chronic stress experience and drives depression onset. *Neuron*, 110:1400–1415.e6, 2022.
- [460] S. Zhu, J.M. Hasse, and F. Briggs. Optogenetic suppression of corticogeniculate feedback in anesthetized ferrets is overridden by visual stimulation. *bioRxiv*, June 2021.
- [461] Yingjie Zhu, Carl FR Wienecke, Gregory Nachtrab, and Xiaoke Chen. A thalamic input to the nucleus accumbens mediates opiate dependence. *Nature*, 530(7589):219–222, 2016.
- [462] Richard E Zigmond, Michael A Schwarzschild, and Ann R Rittenhouse. Acute regulation of tyrosine hydroxylase by nerve activity and by neuro-

- transmitters via phosphorylation. *Annual Review of Neuroscience*, 12(1): 415–461, 1989.
- [463] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2):265–286, 2006.
- [464] Ioannis S Zouridis, Lisa Schmors, Kathrin Maite Fischer, Philipp Berens, Patricia Preston-Ferrer, and Andrea Burgalossi. Juxtacellular recordings from identified neurons in the mouse locus coeruleus. *European Journal of Neuroscience*, 2024.
- [465] Ishan Zutshi, Min Lin Fu, Viranan Lilascharoen, Jill K Leutgeb, Byung K Lim, and Stefan Leutgeb. Recurrent circuits within medial entorhinal cortex superficial layers support grid cell firing. *Nature Communications*, 9(1):3701, 2018.

A | Supplementary material for Study I:  
Schmors, Kotkat, and Bauer et al. (2023)

**A.1 Supplementary Figures**

## A. Supplementary material for Study I: Schmors, Kotkat, and Bauer et al. (2023)

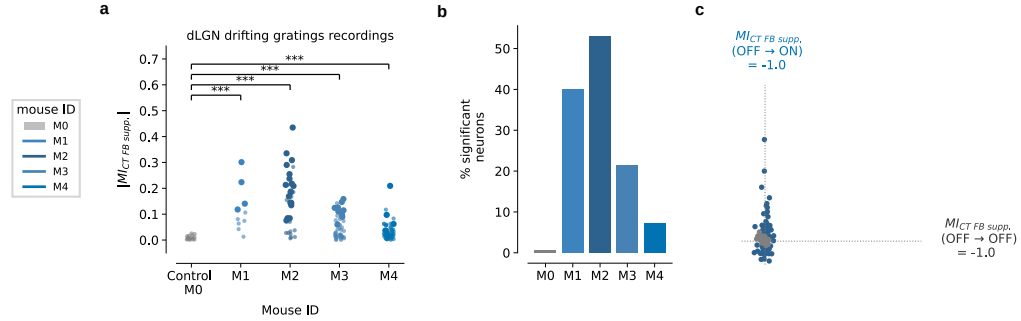


Figure A.1: **Efficiency of optogenetic manipulation of CT feedback.** (a) Optogenetic modulation index (OMI) values for dLGN neurons ( $n = 122$ ), separately by mouse. The four included mice showed significantly higher modulation indices than the control mouse M0 ( $p < 0.01$ , Mann-Whitney-U test). *Opaque dots*: Neurons that are significantly modulated by CT feedback suppression ( $p < 0.05$ , trial-based permutation test). *Transparent dots*: Neurons that did not pass the significance test ( $p > 0.05$ ). (b) Percentage of neurons that were significantly modulated by CT feedback suppression per mouse ( $p < 0.05$ , trial-based permutation test). (c) Scatter of OMI values for the four included mice (*blue*) and the control mouse (*Gray*) for the OFF-ON transitions (i.e., triggered to the onset of CT feedback suppressing light pulses) vs. OFF-OFF transitions (i.e., triggered to times without onset of CT feedback suppressing light pulses).

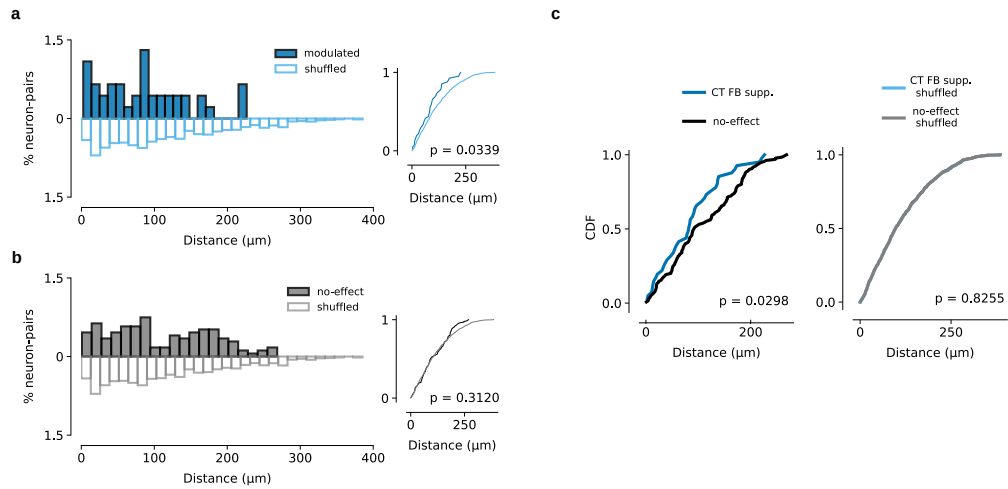


Figure A.2: **CT feedback modulated neurons are closer to each other in the mouse dLGN.** We identified all simultaneously recorded neuron pairs ( $n = 171$ ), of which at least one neuron was positively or negatively modulated by CT feedback suppression ( $|MI_{CT\ FB\ supp.}| \geq 0.1$ ), and calculated the distance in micrometer between each pair on the electrode. (a) *Left, solid bars*: Percentage of neuron pairs ( $n = 41$ ) at each distance from each other on the electrode, when both neurons were modulated by CT feedback suppression. *Left, hollow bars*: Percentage of the same neuron-pairs with shuffled distances. *Right*: cumulative distributions of true and shuffled data ( $p = 0.0339$ , Kolmogorov-Smirnov test). (b) Same as (a), when only one neuron was modulated by CT feedback suppression ( $n = 130$ ).  $p = 0.312$ , Kolmogorov-Smirnov test. (c) *Left*: Cumulative distributions of distances between neuron pairs ( $p = 0.0298$ , Kolmogorov-Smirnov test). *Right*: cumulative distributions of shuffled data ( $p = 0.8255$ , Kolmogorov-Smirnov test), re-plotted from (a) and (b).

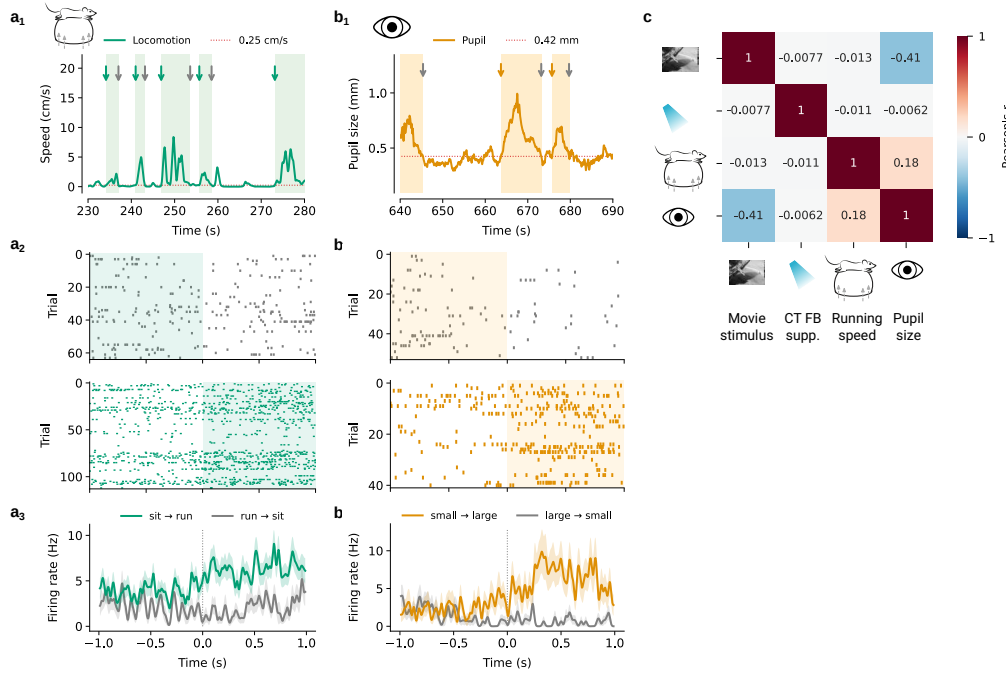


Figure A.3: **CT feedback and behaviour modulates dLGN responses to movie clips.** (a) Modulation of firing rate by running speed for one dLGN example neuron: (a<sub>1</sub>) We considered periods as “running” if the animal’s speed was > 1 cm/s, and “sitting” if the speed was < 0.25 cm/s (red). Transitions between periods are marked by green and grey arrows, respectively. (a<sub>2</sub>) Spike raster plot triggered on transitions from running to sitting (top), and sitting to running (bottom). (a<sub>3</sub>) Corresponding PSTHs. Green: transitions from sitting to running; grey: transitions from running to sitting. Only periods with durations > 2 s were considered. (b) Modulation of firing rate by pupil size for one dLGN example neuron: (b<sub>1</sub>) We considered periods as “large pupil size” when pupil size exceeded the median pupil size during the experiment (red), and “small pupil size” if it was smaller than the 25th percentile. Transitions between periods are marked by orange and grey arrows, respectively. (b<sub>2</sub>) Same as in a<sub>2</sub> but for transitions from large to small pupil (top) and from small to large pupil (bottom). (b<sub>3</sub>) Same as in a<sub>3</sub> but for pupil size transitions. (c) Correlation matrix of average stimulus intensity, optogenetic light pulses, running speed and pupil size traces. Values denote average correlation values across experiments ( $n = 10$ ).

A. Supplementary material for Study I:  
Schmors, Kotkat, and Bauer et al. (2023)

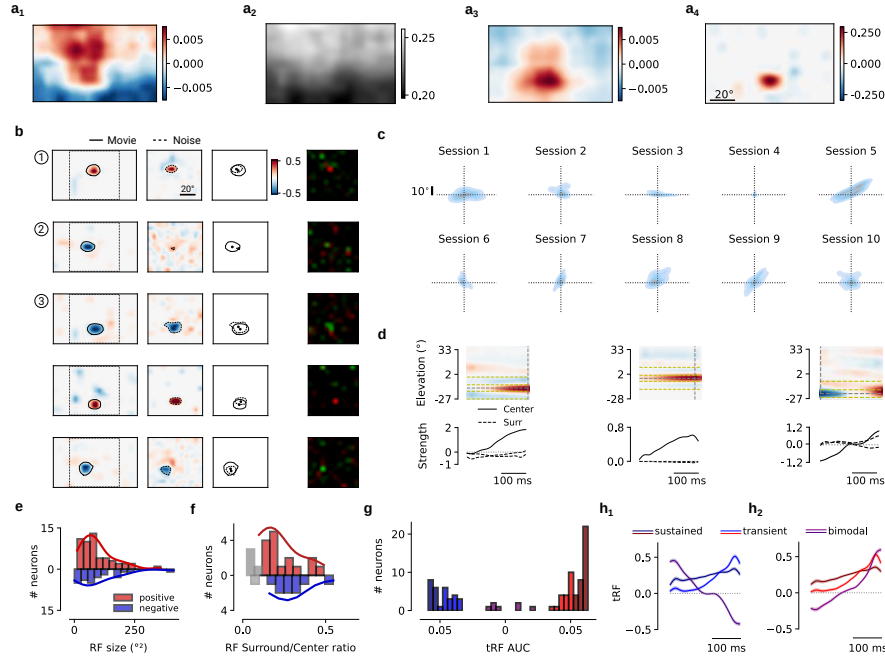


Figure A.4: **Spatial RF estimation by the spline-GLM model.** (a) Comparison of spatial RF (sRF) estimates during the movie stimulus for one example dLGN neuron. (a1) Simple spike triggered average (STA). (a2) Pixel intensity averaged across all movie frames, showing a gradient from top to bottom typical for naturalistic scenes. (a3) STA corrected for average intensity gradient. This yields a RF with similar centre and polarity compared to the one learned by the spline-GLM (a4), but with larger and more diffuse area. (a4) Learned sRF by the spline-GLM model. (b) To validate the sRF on a neuron-by-neuron basis, we compared the spline-GLM sRF learned from movies (*first column*) with spline-GLM sRF learned from sparse noise experiments (*second column*; overlay of the two sRF centres and contours in the *third column*; see Methods) and with the stimulus-triggered average response map for the sparse noise experiments computed independently from the model (*last column*). The first three example neurons are the same as in ??c. (c) Centre shift between movie and sparse noise sRFs for all 10 experimental sessions (neurons with significant stimulus kernels,  $n = 85$ ). Position around the origin indicates close to zero shift. (d) Spatio-temporal RF (stRF) for three example neurons. *Top*: stRFs collapsed across the azimuth axis, and plotted as a function of elevation and time. *Vertical dashed line*: time point with peak activity. *Horizontal dashed lines*: centre regions defined according to the spatial width of the peak activity, and surround regions defined as a ring encircling the centre and extending up to trice the diameter or  $9^\circ$ . *Bottom*: stRF centre (*solid*) and surround (*dashed*) activity. *Left*: antagonistic surround neuron (surround-to-centre ratio = 0.32). *Middle*: non-antagonistic surround neuron (surround-to-centre ratio = 0.02). *right*: bi-phasic response neuron. (e) Area of sRF centres, separately for neurons with positive (*red*) and negative (*blue*) RF centre and significant stimulus kernels. Of these,  $\sim 61\%$  (52/85) had a positive centre,  $\sim 29\%$  (25/82) had a negative centre, and  $\sim 9\%$  (8/85) had a bimodal response where the detected RF changed its polarity across time. Median RF radius: 5 deg; median centre area:  $81 \text{ deg}^2$  for positive RF centers and  $77 \text{ deg}^2$  for negative RF centers. (f) Surround-to-centre ratios for the population of dLGN neurons with significant stimulus kernels, and antagonistic activity in the surround region ( $\sim 41\%$  (35/85) neurons). We defined a threshold at 0.1 surround-to-centre ratio (*gray*). Of the remaining neurons with pronounced antagonistic surround, 16 neurons had positive RF centre with a negative surround (*red*), 9 neurons had negative RF centre with a positive surround (*blue*), and 4 neurons showed a biphasic response together with an antagonistic surround. (g) Quantification of tRF properties: histogram of area under the tRF, separately for positive (*red*) and negative (*blue*) RF centre. Neurons with small tRF AUC values 0.035 represent the bimodal neurons ( $n = 8$ , *purple*). Neurons with the largest tRF AUC values  $\geq 0.052$  were defined as neurons with sustained responses, while the remaining neurons were defined as neurons with transient responses. (h) Mean tRF split into sustained, transient, and biphasic.

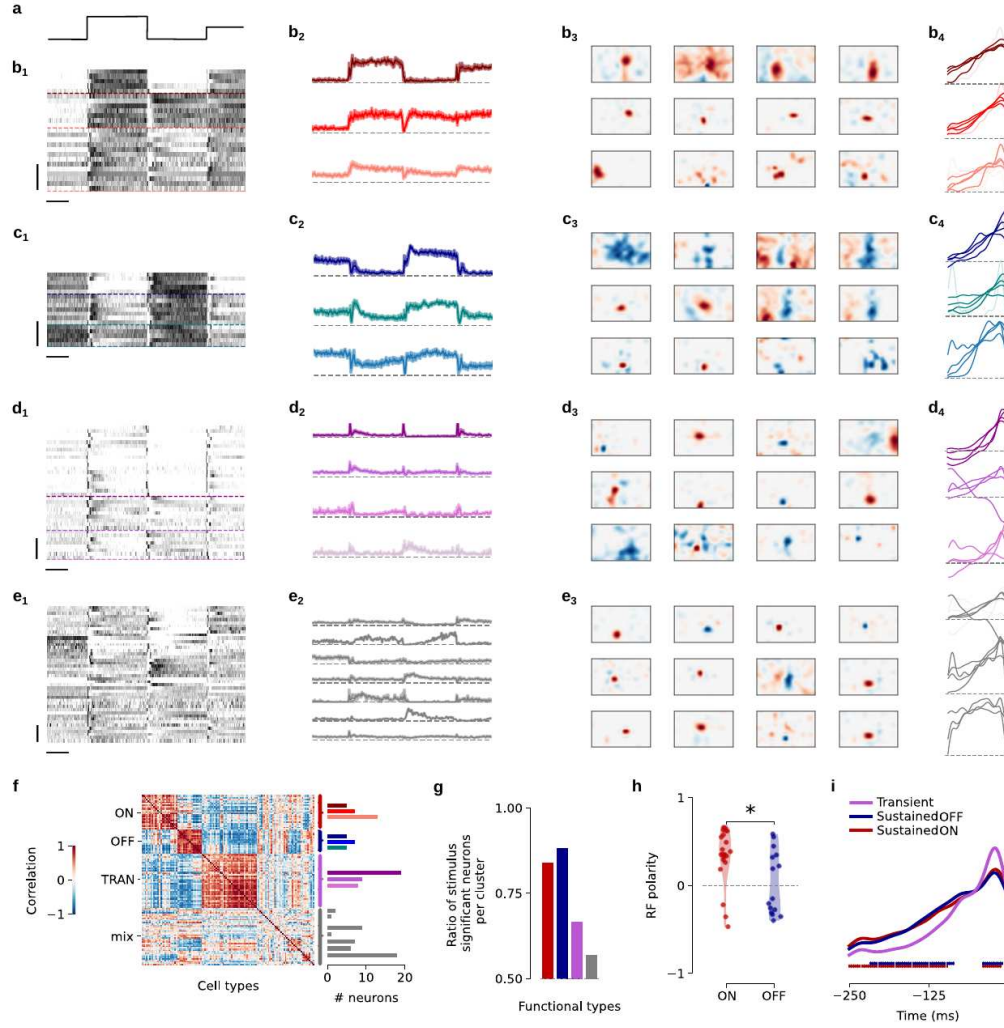


Figure A.5: **Spline-based GLM captures spatio-temporal RFs of different functional cell types in dLGN.** (a) *Top, left:* Light intensity steps of the full field stimulus used for functional cell type classification. (b) Responses of all neurons assigned to the Sustained ON group ( $b_1$ ,  $n = 25$ ), along with their average sub-cluster responses ( $b_2$ ). Spatial ( $b_3$ ) and temporal RF kernels ( $b_4$ ) of example neurons with significant movie stimulus kernel learned by the spline-GLM independently of the clustering. (c) Same as (b), for the Sustained OFF group ( $n = 17$ ). (d) Same as (b), for the Transient group ( $n = 36$ ). (e) Same as (b), for the remaining groups (mixed cluster,  $n = 44$ ). (f) *Left:* Correlation matrix of PCA features used for clustering. Neurons are sorted by groups. *Right:* Neuron counts per group. (g) Fraction of neurons with significant stimulus kernel in the spline-GLM model per functional type based on clustering. (h) Spatial RF polarity (sign of the spline-GLM fitted RF centre) for the ON and the OFF group based on clustering ( $p = 0.018$ , Mann-Whitney-U test). (i) Average temporal RF of movie kernels learned from spline-GLM for the OFF, ON, and Transient groups. *Horizontal bars:* Significant differences between OFF and Transient groups (blue) and between ON and Transient groups (red).

A. Supplementary material for Study I:  
Schmors, Kotkat, and Bauer et al. (2023)

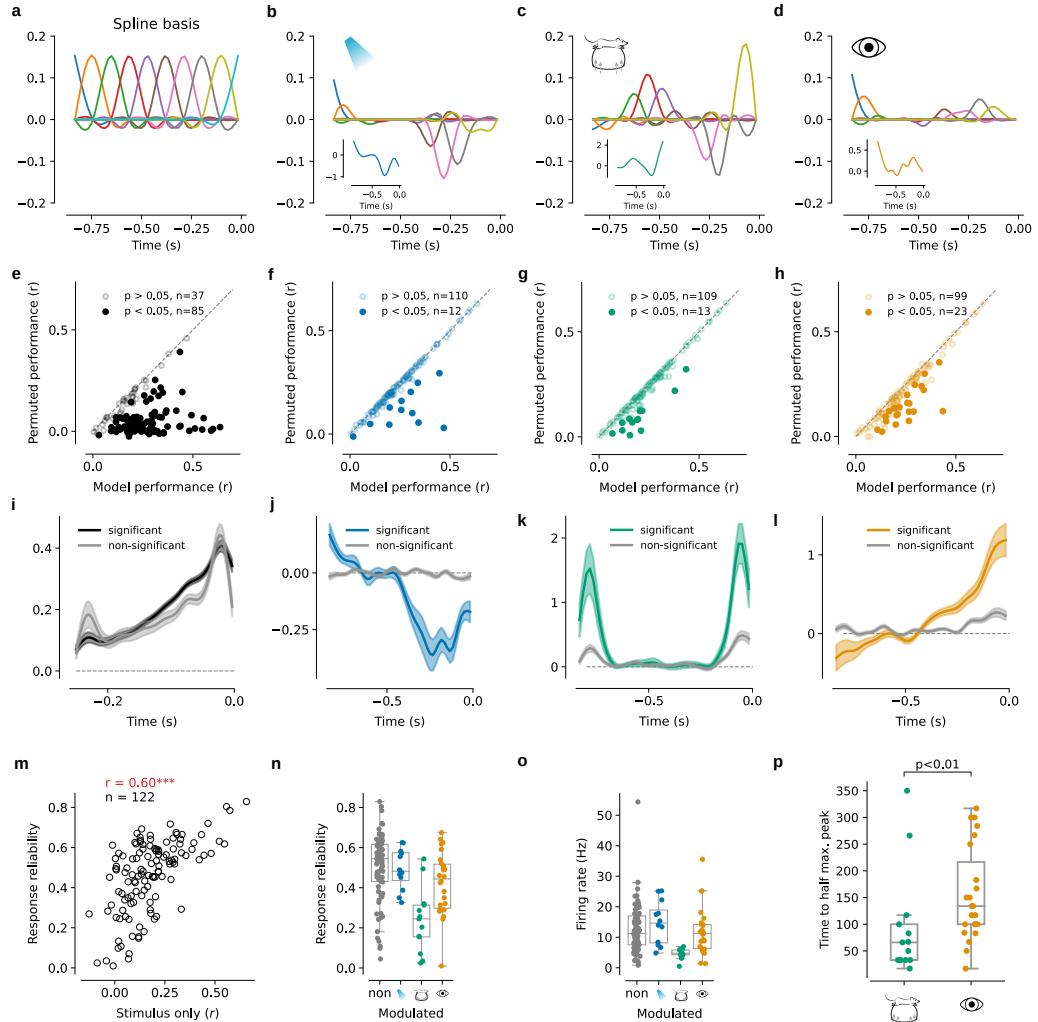


Figure A.6: **Spline-GLM model kernels for modulatory inputs.** (a) Illustration of the natural cubic spline basis functions ( $n=10$ ). (b) Weighted bases of the natural cubic splines after fitting the CT feedback kernel to the data for one example neuron. CT feedback kernel is shown in the inset. (c) Same as (b), for run speed kernel. (d) Same as (b), for pupil size kernel. (e) Model performance for permuted visual stimulus input (mean over all permutations) vs. actual performance. Filled circles: significant neurons. (f) Same as (a), for CT feedback suppression. (g) Same as (a), for running. (h) Same as (a), for pupil size. (i) Average temporal RF kernels, separately for neurons with significant (black,  $n = 85$ ) versus non-significant (grey,  $n = 37$ ) spatio-temporal kernels. Shaded areas: SEM. (j) Same as (e), for neurons with significant (blue,  $n = 12$ ) versus non-significant (grey,  $n = 110$ ) CT feedback kernels. (k) Same as (e), for neurons with significant (green,  $n = 13$ ) versus non-significant (grey,  $n = 109$ ) run kernels. (l) Same as (e), for neurons with significant (orange,  $n = 23$ ) versus non-significant (grey,  $n = 99$ ) pupil kernels. (m) Model performance versus response reliability across all neurons. Response reliability was computed as the mean correlation of dLGN responses to each 40 s repeated movie clip with all other 8 repeats. (n) Response reliability of dLGN neurons split into neurons with significant kernels to CT feedback suppression, running, pupil size, or non-modulated. (o) Same as (j) but for firing rate of dLGN neurons. (p) Kernel time to half maximum peak for run and pupil size kernels.

A. Supplementary material for Study I:  
Schmors, Kotkat, and Bauer et al. (2023)

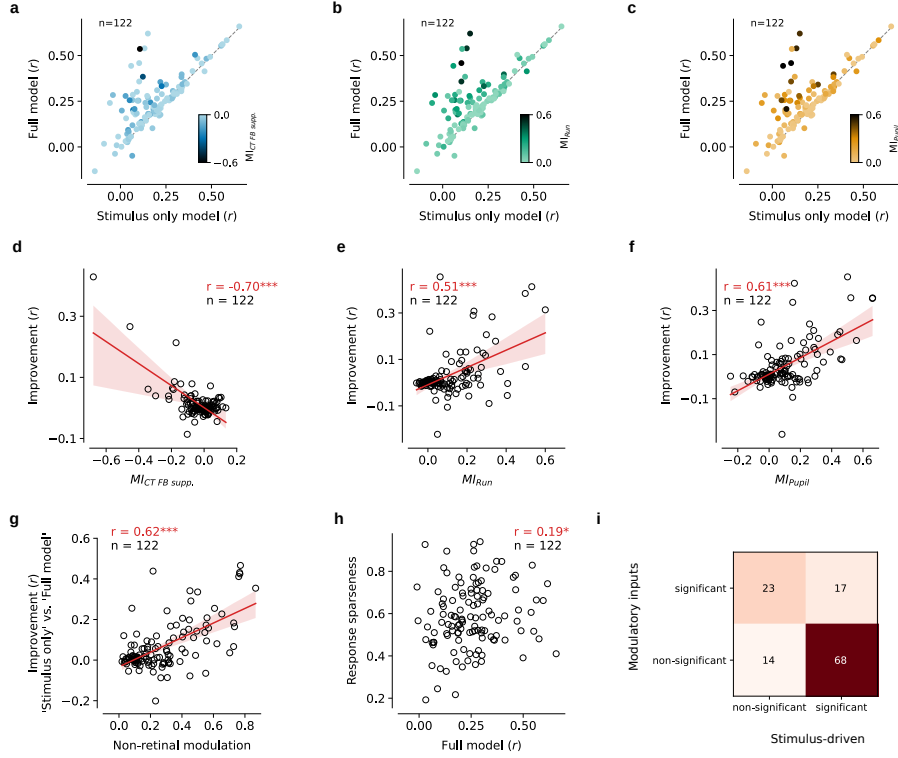


**Figure A.7: Model performance depends on various factors.** (a) Comparison of model performances on the test set for all neurons in the 'Stimulus only' model and the 'Full model'. Darker colours indicate stronger modulation by CT feedback quantified by  $MI_{CT\ FB\ supp.}$  (see Methods). (b) Same as in (a) but for by running quantified by the modulation index  $MI_{Run}$ . (c) Same as in (a) but for pupil size quantified by the modulation index  $MI_{Pupil}$ . (d) Correlation of modulation by CT feedback suppression ( $MI_{CT\ FB\ supp.}$ ) and model performance improvement. To measure improvement two models were compared: the 'Stimulus only' model and a model with Stimulus and CT feedback suppression as inputs. (e) Same as in h but for modulation by running ( $MI_{Run}$ ). The 'Stimulus only' model was compared with a model that integrated stimulus and running as inputs. (f) Same as in h but for modulation by pupil size ( $MI_{Pupil}$ ). The 'Stimulus only' model was compared with a model that integrated stimulus and pupil size as inputs. (g) Modulation by additional factors ( $MI_{Joint}$ ) versus improvement of model performance when comparing the 'Stimulus only' model and the 'Full model'.  $***p \leq 0.001$  (h) Model performance versus response sparseness. Sparseness was computed according to [417].  $*p \leq 0.05$  (i) The matrix illustrates the neuron counts categorised into 68 neurons significant only to feedforward stimulus-related predictor, 23 neurons significant only to one or more of the predictors for modulatory inputs (CT feedback, run speed, and pupil dilation), 17 significant to all predictors, and 14 not exhibiting significance for any of the predictors.

A. Supplementary material for Study I:  
Schmors, Kotkat, and Bauer et al. (2023)

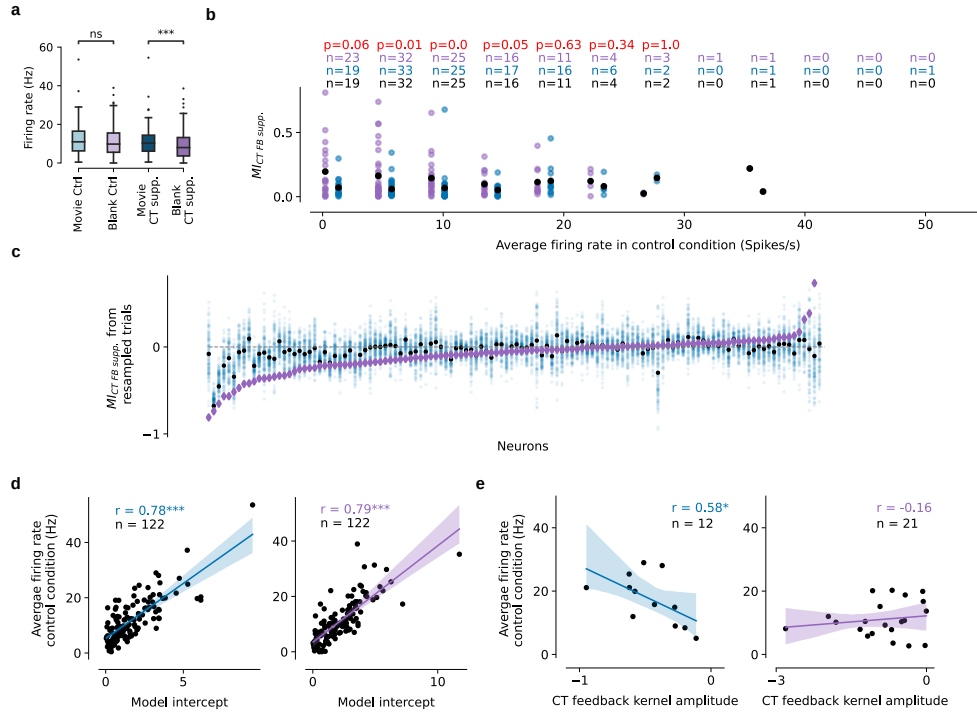


**Figure A.8: Influences of visual stimulus on effects of CT FB suppression in dLGN.** (a) Firing rate comparison between conditions. Bonferroni corrected  $p$ -values of paired Wilcoxon signed-rank test:  $***p \leq 1.0 \times 10^{-3}$ ; 'ns' non-significant. (b) Analysis controlling potential influences of overall firing rate differences between conditions.  $MI_{CT\ FB\ supp.}$  for movies (blue) or blank periods (purple) as a function of binned average firing rate during the control condition. We counted the neurons that fell within each firing rate bin for the two stimulus conditions (top, coloured numbers). To ensure equal number of neurons per firing rate bin between the conditions, we randomly sampled the minimum number of neurons for that bin for one of the stimulus conditions (black numbers). Black dots represent the mean  $MI_{CT\ FB\ supp.}$  value per bin per condition after potential re-sampling. Even after approximating overall firing rates,  $MI_{CT\ FB\ supp.}$  remained significantly higher for low firing rate bins, where most neurons were sampled (red numbers, Mann-Whitney-U test). (c) Scatters of  $MI_{CT\ FB\ supp.}$  values recalculated based on re-sampled trials. Given that movie and blank conditions had different number of trials, and in order to rule out that the number of trials might affect the  $MI_{CT\ FB\ supp.}$  values and accordingly the reported results, we performed 100 repeats of a re-sampling procedure. During each repeat, we randomly sampled 24 trials for each neuron during movie condition and re-calculated the  $MI_{CT\ FB\ supp.}$  values (light blue dots). We found that the original movie  $MI_{CT\ FB\ supp.}$  values (black dots) well represented the median of the re-calculated values. We also found that our observation that the original Blank  $MI_{CT\ FB\ supp.}$  (purple) being stronger than movie  $MI_{CT\ FB\ supp.}$  held true after re-sampling. (d) Scatter plot of average firing rate during the control condition vs. intercept fitted by the model for the movie (left) and blank periods (right). Panels (a–d) show data from all 122 neurons. (e) Scatter plot of average firing rate vs. maximum amplitude of the fitted CT feedback kernels for significantly modulated neurons for the movie (left) and blank (right) conditions.

## A.2 Materials and methods

All procedures complied with the European Communities Council Directive 2010/63/EU and the German Law for Protection of Animals, and were approved by local authorities, following appropriate ethics review.

### A.2.1 Surgical procedures

Experiments were carried out under under Licence ROB-55.2-2532.Vet\_02-17-40 in 4 adult *Ntsr1-Cre* mice (median age:  $15.5 \pm 6.45$  weeks; B6.FVB(Cg)-Tg(*Ntsr1-cre*)GN220 *Gsat/Mmc*d; MMRR) of either sex. Stereotactic surgeries were performed to implant a head-post for head-fixation, implant a ground/reference screw for electrophysiology, inject a virus for optogenetic feedback manipulation, and drill a craniotomy for acute electrode insertions.

#### Stereotactic surgery preparation and initiation

Thirty minutes prior to the surgical procedure, mice were injected with an analgesic (Metamizole, 200 mg/kg, sc, MSD Animal Health, Brussels, Belgium). To induce anesthesia, animals were placed in an induction chamber and exposed to isoflurane (5% in oxygen, CP-Pharma, Burgdorf, Germany). After induction of anesthesia, mice were fixated in a stereotaxic frame (Drill & Microinjection Robot, Neurostar, Tuebingen, Germany) and the isoflurane level was lowered (0.5 %–2 % in oxygen), such that a stable level of anesthesia could be achieved as judged by the absence of an interstitial reflex. Throughout the procedure, the eyes were covered with an eye ointment (Bepanthen, Bayer, Leverkusen, Germany) and a closed loop temperature control system (ATC 1000, WPI Germany, Berlin, Germany) ensured that the animal's body temperature was maintained at 37 °C. At the beginning of the surgical procedure, an additional analgesic was administered (Buprenorphine, 0.1 mg/kg, sc, Bayer, Leverkusen, Germany) and the animal's head was shaved and thoroughly disinfected using iodine solution (Braun, Melsungen, Germany). Before performing a scalp incision along the midline, a local analgesic was delivered (Lidocaine hydrochloride, sc, bela-pharm, Vechta, Germany). The skin covering the skull was partially removed and cleaned from tissue residues with a drop of H<sub>2</sub>O<sub>2</sub> (3 %, AppliChem, Darmstadt, Germany). Using four reference points (bregma, lambda, and two points 2 mm to the left and to the right of the midline respectively), the animal's head was positioned into a skull-flat configuration for the further steps.

#### Virus injection

In order to suppress V1 L6 CT FB selectively and reversibly, we conditionally expressed the chloride-conducting channelrhodopsin *stGtACR2* [142, 241, 436] in L6a CT pyramidal cells [52, 362, 367] by injecting AAV-*stGtACR2*-RFP into the left hemisphere V1 of *Ntsr1-Cre* mice [140, 187, 238] (Figure 3.1a). *Ntsr1+* neurons are known to correspond with > 90% specificity to L6 CT pyramidal cells [46, 127, 291]. Furthermore, the opsin *stGtACR2* restricts expression to somata and the axon-initial segment which prevents possible accidental axonal depolarization due to a differential Cl<sup>-</sup> ion reversal potential across different neuronal compartments [240, 434, 435]. It also offers improved photocurrents and higher sensitivity, which are of particular relevance to manipulating deeply located L6 CT neurons, while avoiding light artifacts and tissue damage arising from excessive light intensities [435].

## A. Supplementary material for Study I: Schmors, Kotkat, and Bauer et al. (2023)

---

Before surgery, the Cre-dependent, stGtACR2-expressing *adeno-associated virus* (AAV) vector (pAAV\_hSyn1-SIO-stGtACR2-FusionRed, Addgene, #105677) stock solution was diluted to  $5 \times 10^{11}$  gc/ml titers, and aliquotted to 4  $\mu$ L.

During surgery, aliquots were front-loaded into a glass pipette mounted on a Hamilton syringe (SYR 10  $\mu$ L 1701 RN no NDL, Hamilton, Bonaduz, Switzerland), controlled by the Injection Robot of the Neurostar Stereotax. After performing a small craniotomy for injection (100  $\mu$ m diameter), we injected 300 nl of virus solution into V1 (2 $\times$ 50 nl shots injected at a rate of 50 nl / 30 s at a respective depth of 900  $\mu$ m, 800  $\mu$ m and 700  $\mu$ m below the brain surface.

### Head-post and ground and reference screw implantation

For implant fixation, the exposed skull was covered with OptiBond FL primer and adhesive (Kerr Dental, Rastatt, Germany) omitting three locations: V1 (AP:  $-3.28$  mm, ML:  $-2.4$  mm), dLGN (AP:  $-2.3$  mm, ML:  $-2$  mm), and a position roughly 1.5 mm anterior and 1 mm to the right of bregma, designated for a miniature ground and reference screw.

A custom-made lightweight stainless steel head bar was positioned over the posterior part of the skull such that the round opening in the bar was centered on V1/dLGN. The head bar was attached with dental cement (Ivoclar Vivadent, Ellwangen, Germany) to the primer/adhesive. The opening was later filled with the silicone elastomer sealant Kwik-Cast (WPI Germany, Berlin, Germany). Then the miniature screw (00-96 X 1/16 stainless steel screws, Bilaney), which served both as ground and reference that was soldered to a custom-made connector pin, was implanted.

### Post-surgical treatment and animal setup habituation

At the end of the procedure, an iodine-based ointment (Braunodivon, 10%, B. Braun, Melsungen, Germany) was applied to the edges of the wound and a long-term analgesic (Meloxicam, 2 mg/kg, sc, Böhringer Ingelheim, Ingelheim, Germany) was administered and for 3 consecutive days. For at least 5 days post-surgery, the animal's health status was assessed via a score sheet.

After at least 1 week of recovery, animals were gradually habituated to the experimental setup by first handling them and then simulating the experimental procedure. To allow for virus expression, neural recordings started after an incubation time of 2-4 weeks after injection.

### Craniotomy

On the day prior to the first day of recording, mice were fully anesthetized using the same procedures as described for the initial surgery, and a craniotomy (ca. 2 $\times$ 1 mm on the AP $\times$ BL axes) was performed over dLGN (ca. 2.5 mm posterior from bregma and 2.3 mm lateral from midline) and V1 and re-sealed with Kwik-Cast (WPI Germany, Berlin, Germany). As long as the animals did not show signs of discomfort, the long-term analgesic Metacam was administered only once at the end of surgery, to avoid any confounding effect on experimental results. Recordings were performed daily and continued for as long as the quality of the electrophysiological signals remained high.

### A.2.2 Extracellular multi-electrode array (MEA) recordings

After 2-4 weeks of incubation time, we performed *in vivo* extracellular multi-electrode array (MEA) recordings of dLGN neurons in awake, head-fixed mice (**Figure 3.1a**). Extracellular signals were recorded at 30 kHz (Blackrock microsystems, Blackrock Microsystems Europe GmbH, Hanover, Germany). For each recording session, the silicon plug sealing the craniotomy was removed. For dLGN recordings, a 32 channel linear silicon probe (Neuronexus A1x32Edge-5mm-20-177-A32) was lowered to a depth of  $\sim 2500$ – $3500 \mu\text{m}$  below the brain surface. We judged recording sites to be located in dLGN based on the characteristic progression of RFs from upper to lower visual field along the electrode shank [313], the presence of responses strongly modulated at the temporal frequency of the drifting gratings (F1 response), and the preference of responses to high temporal frequencies [145, 313]. For *post hoc* histological reconstruction of the recording site, the electrode was stained with DiI (Invitrogen, Carlsbad, USA) for some (typically the last) recording sessions.

### A.2.3 Locomotion

During the experiment, mice were free to run on an air-floating Styrofoam ball and the run speed was recorded via locomotion sensors (**Figure 3.1a**). Two optical computer mice interfaced with a microcontroller (Arduino Duemilanove) sampled ball movements at 90 Hz.

To compute animal run speed, we used the Euclidean norm of three perpendicular components of ball velocity (roll, pitch and yaw) and smoothed traces with a Gaussian kernel ( $\sigma = 0.2 \text{ s}$ ). To quantify the effect of running vs. sitting on various response properties, the run modulation index ( $MI_{Run}$ ) was defined based on the mean firing rates during running vs. sitting periods as

$$MI_{Run} = \frac{\text{running} - \text{sitting}}{\text{running} + \text{sitting}}, \quad (\text{A.1})$$

where running periods were defined as those for which speed exceeded 1 cm/s, and sit periods as those for which speed fell below 0.25 cm/s.

To test for a significant difference in mean FRs between the run vs. sit conditions matched for each neuron, we used the Wilcoxon signed-rank test.

### A.2.4 Eye tracking

To record eye position and pupil size, the animal's eye that was viewing the stimulus was illuminated with infrared LED light and monitored using a zoom lens (Navitar Zoom 6000) coupled with a camera (Guppy AVT camera; frame rate 50 Hz, Allied Vision, Exton, USA).

Pupil position was extracted from the eye-tracking videos using a custom, semi-automated algorithm. Briefly, each video frame was equalized using an adaptive bi-histogram equalization procedure, and then smoothed using median and bilateral kernels. The center of the pupil was detected by taking the darkest point in a convolution of the kernelled image with a black square. Next, the peaks of the image gradient along lines extending radially from the center point were used to define the pupil contour. Lastly, an ellipse was fit to the contour, and the center and area of this ellipse was taken as the position and size of the pupil, respectively. A similar procedure was used to extract the position of the corneal reflection (CR) of the LED illumination. Eye-closure, grooming, or implausible ellipse fitting was automatically detected and the adjacent data points 0.15 s before and after

were excluded. Linear interpolation and a subsequent Gaussian smoothing ( $\sigma = 0.06$  s) was applied to fill the removed segments. Adjustable algorithm parameters, such as the threshold of the mean pixel-wise difference between each frame and a reference frame to detect blinks, were set manually for each experiment.

To quantify the effect of large vs. small pupil sizes on various response properties, the eye modulation index ( $MI_{Pupil}$ ) was defined based on the mean firing rates during periods of large vs. small pupils as

$$MI_{Pupil} = \frac{\text{pupil large} - \text{pupil small}}{\text{pupil large} + \text{pupil small}}, \quad (\text{A.2})$$

where periods of large pupils were defined as those for which pupil size was above the 50<sup>th</sup> percentile of the median normalized pupil trace, and periods of small pupils as those for which pupil size fell below the 25<sup>th</sup> percentile.

### A.2.5 Optogenetic feedback suppression

To photosuppress V1 Ntsr1+ L6 CT pyramidal cells, an optic fiber (480  $\mu\text{m}$  core diameter, MFP\_480/500/1000-0.63\_m\_SMA, Doric Lenses, Quebec, Canada) was coupled to a light-emitting diode (blue LED, center wavelength 465 nm, LEDC2\_465/635\_SMA, Doric Lenses, Quebec, Canada) and positioned with a micromanipulator less than 1 mm above the exposed surface of V1. A black metal foil surrounding the tip of the head bar holder prevented the photostimulation light from reaching the animal's eyes. To ensure that the photostimulation was effective, the first recording session for each mouse was carried out in V1. Only if the exposure to light reliably induced suppression of V1 activity was the animal used for subsequent dLGN recordings. LED light intensity was adjusted on a daily basis to evoke reliable effects and account for variations in exact virus titer, volume, incubation time, virus expression levels, and fiber position (0.85-9.5 mW at the fiber tip). Since the tip of the fiber never directly touched the surface of the brain, and since the clarity of the surface of the brain varied (generally decreasing every day following the craniotomy), the light intensity delivered even to superficial layers of V1 was inevitably lower. For the movie stimulus, optogenetic pulses of 1 s duration were sent randomly each second with a 50 % chance.

To quantify the effect of CT feedback suppression on various response properties, we defined the optogenetic modulation index ( $MI_{CT\ FB\ supp.}$ ) based on the mean FRs during CT feedback suppression ('opto') versus the control condition as

$$MI_{CT\ FB\ supp.} = \frac{\text{opto} - \text{control}}{\text{opto} + \text{control}} \quad (\text{A.3})$$

### A.2.6 Joint modulation by additional inputs

In order to quantify the joint effect of all the modulatory inputs, we calculated  $MI_{Joint}$  as

$$MI_{Joint} = 1 - (1 - MI_{CT\ FB\ supp.})(1 - MI_{Run})(1 - MI_{Pupil}) \quad (\text{A.4})$$

### A.2.7 Visual stimulation

During the experiment, the mice were passively viewing visual stimuli on an LCD monitor screen in their right visual field. The visual stimuli were presented on a gamma-calibrated

liquid crystal display (LCD) monitor (Samsung SyncMaster 2233RZ,  $47 \times 29$  cm,  $1680 \times 1050$  resolution at 60 Hz, mean luminance  $50 \text{ cd/m}^2$ ) positioned at a distance of 25 cm from the animal's right eye (spanning  $\sim 108 \times 66$  visual angle by small angle approximation) using custom written software (EXPO, <https://sites.google.com/a/nyu.edu/expo/home>). The display was gamma-corrected for the presentation of artificial stimuli, but not for movies (see below).

### Movie stimulus

For movie stimulus generation, we adopted a set of randomly picked clips from various movies. Briefly, source movie clips were converted to grey scale, temporally downsampled to 30 frames per s, spatially resampled and cropped to  $424 \times 264$  pixels, to be presented on our  $47 \times 29$  cm monitor screen at 25 cm distance at  $106 \times 66$  (4 pixels/) visual angle (by small angle approximation, which preserves the desired pixel resolution at the screen center better than the arctangent). Movie frames were not histogram-equalized and presented at 60 Hz (repeating each frame twice) without monitor gamma correction, since cameras are already gamma corrected for consumer displays [319]. To generate the movie sequence, we used a random set of 296 unique movie clips (5 s each) and split 188/296 clips into 8 parts of 36 unique clips ( $5 \times 36 = 180$  s per part). They were interleaved with set of 8 clips ( $5 \times 8 = 40$  s) which was repeated 9 times. The repeated clips served to give an estimate of response reliability to the same clips. The movie sequence was flanked by a period of blank grey screen presentation (1 min) at the beginning and at the end, to record spontaneous activity. This resulted in a total stimulus duration of  $\sim 32$  mins. To rule out sequence effects, we randomized the clip order for different stimulus presentations. To investigate the effects of L6 CT FB suppression, we simultaneously presented a random optogenetic pulse train of 1 s pulses, occurring each second with a probability of 50 %, throughout the entire stimulus duration, including blank grey screen periods.

### Sparse noise stimulus

To measure RFs in a more standard manner, we also presented an (artificial) sparse noise stimulus. The stimulus consisted of a rapid sequence of non-overlapping white and black squares appearing in succession within a  $12 \times 12$  square grid presented on a grey background of mean luminance ( $50 \text{ cd/m}^2$ ). The square grid spanned 60 per side, while individual squares spanned 5 per side. Each square flashed 20 times for 200 ms at random order. The stimulus triggered average (STA) for the sparse noise stimulus was computed using the onset of each square and then computing the normalised mean spike rate triggered by each position.

## A.2.8 Histology

To verify virus expression and recording sites, we performed post-mortem histological analyses. After the final recording session, mice were first administered an analgesic (Metamizole, 200 mg/kg, sc, MSD Animal Health, Brussels, Belgium) and following a 30 min latency period were transcardially perfused under deep anesthesia using a cocktail of Medetomidin (Domitor, 0.5 mg/kg, Vetoquinol, Ismaning, Germany), Midazolam (Climasol, 5 mg/kg, Ratiopharm, Ulm, Germany) and Fentanyl (Fentadon, 0.05 mg/kg, Dechra Veterinary Products Deutschland, Aulendorf, Germany) (ip). Perfusion was first done with Ringer's lactate solution followed by 4% paraformaldehyde (PFA) in 0.2 M sodium phosphate buffer (PBS). Brains were removed, postfixed in PFA for 24 h, and then rinsed with and stored in PBS at 4 C. Slices ( $50 \mu\text{m}$ ) were cut using a vibrotome (Leica VT1200 S, Leica, Wetzlar, Germany), stained with DAPI-solution (DAPI, Thermo

Fisher Scientific, Waltham, Massachusetts, USA), mounted on glass slides with Vectashield mounting medium (Vectashield H-1000, Vector Laboratories, Burlingame, USA), and coverslipped. A scanning fluorescent microscope (BX61, Olympus, Tokyo, Japan) was used to inspect slices for the presence of red fluorescent protein (RFP/FusionRed) marking stGtACR2-channels, and DiI, marking electrode tracks. Recorded images were processed off-line using FIJI [330, 345].

### A.2.9 Control experiments with Ntsr-negative mouse

Additionally, same experimental procedures were followed for control (Ntsr-negative) mouse, in order to assess the efficiency of optogenetic manipulations of CT feedback. Prior to experimentation, the genotype of the mice was confirmed via polymerase chain reaction (PCR) analysis.

### A.2.10 Spike sorting and unit extraction

Spike sorting was performed to obtain single unit activity from extracellular recordings. Electrophysiological signal recordings were kernalized using a 4<sup>th</sup>-order Butterworth high-pass non-causal kernel with a low frequency cutoff of 300 Hz. We then used the open source, MATLAB-based (The Mathworks, Natick, Massachusetts, USA), automated spike sorting toolbox Kilosort and Kilosort2 [295]. Resulting clusters were manually refined using Spyke [376], a Python application that allows for the selection of channels and time ranges around clustered spikes for realignment, as well as representation in 3D space using dimension reduction (multichannel PCA, ICA, and/or spike time). In 3D, clusters were then further split via a gradient-ascent based clustering algorithm (GAC) [393]. Exhaustive pairwise comparisons of similar clusters allowed the merger of potentially over-clustered units. For subsequent analyses, we inspected autocorrelograms and mean voltage traces, and only considered units that displayed a clear refractory period and a distinct spike waveshape.

### A.2.11 Data analysis

Data analysis was performed using custom-written code that applies general tools such as Numpy, sklearn, matplotlib, seaborn, pandas and carried out in a MySQL-based database using the DataJoint [450]. We also used a customized version of RFEST (<https://github.com/berenslab/RFEST>) to allow for multiple model inputs.

### Firing rate calculations

To obtain units firing rates in spikes per second (Hz), each unit's spike density function (SDF) was calculated by binning spikes into a firing rate histogram (bin width = 20 ms) and convolving this with a Gaussian of width  $2\sigma = 10$  ms. Mean firing rates (FRs) over a given condition were calculated as the mean of the time-varying firing rates for the defined periods.

### Identification of putative neuronal types

In order to identify putative excitatory and inhibitory neurons in V1, we analysed the extracellular spike waveform. For each neuron, the mean waveform of the maximally

responsive electrode channel was obtained, and the time between trough and peak (trough-to-peak time) and the full-width at half-height of the peak (peak width) were calculated. Using the waveforms of all V1 neurons, a k-means algorithm was used to cluster the data into 2 populations.

## Response reliability

To compute how reliable a neurons responded to the visual stimulus we used the set of 8 clips that where repeated 9 times throughout the experiment. We computed reliability by correlating each repetition with the mean of all other repetitions and averaging that over all splits.

## Optogenetic significance

To compute the reliability of the optogenetic manipulation of L6 CT neurons, we developed a trial-based permutation test. In response to drifting gratings, we calculated the average firing rate during each trial and separated the trial with optogenetic CT feedback suppression ( $trials_{CT\ FB\ supp.}$ ;  $n=90 - -130$ ) from those without ( $trials_{control}$ ;  $n=90 - -130$ ). We calculated the observed statistics  $effect_{CT\ FB\ supp.}$  as the difference in the means between the  $trials_{CT\ FB\ supp.}$  and the  $trials_{control}$ . We assessed significance by permuting trial labels 1000 times and considered the effect significant if it fell outside of the distribution of permuted  $effect_{CT\ FB\ supp.}$ s. Finally, we compared the percentage of neurons that passes the significance test between the included recording and a recording from a control mouse (Ntsr-negative).

## Exclusion criteria

Neurons with mean evoked firing rates  $< 0.1$  Hz were excluded from all further analysis. Further analysis-specific selection criteria are stated in the appropriate subsections.

## Predictor correlations

To test for the correlations between the predictors *stimulus*, *opto*, *run*, and *eye*, we temporally aligned these traces, including only time points for which we had data points in all traces (e.g. removing periods of eye blinks). We then explored their cross-correlations in order to detect potential delays in their effects on each other. Such is the case, for instance in the pupil light reflex, where increases in stimulus intensity are followed by a delayed decrease in pupil size. We then used the delay time to shift the traces appropriately before computing their correlation value (Pearson's  $r$ ).

In order to test for statistical significance of the obtained correlation values, we first needed to account for the fact that our time-series inherently contain autocorrelations which would lead to an overestimation of correlations between them (except for the random opto pulses) [149]. We therefore used a permutation test in which we randomly permuted the stimulus traces for  $k = 1000$  iterations, and then computed the p-value of the observed correlation value as its percentile within the null distribution of p-values for the permuted traces [149].

### A.2.12 Spline-based generalized linear model

To estimate the spatio-temporal RFs (STRFs), we used the RFest Python toolbox for spline-based spatio-temporal RF estimation [174]. Here, the spline-based GLM reduces the number of parameters compared to traditional approaches, which need to estimate every pixel in the RF independently, while spline-based GLM only estimate the weight of the bases. RFest is also less data demanding and reduces the computation time significantly [174]. The number of parameters are given by the number of basis functions, also referred to the degrees of freedom. By using natural cubic splines as the basis, the estimates are automatically smooth, which is a desirable property for single STRFs. To impose sparsity on the weights (also a desirable property of STRFs) we added L1 regularization, which pushed the weights for less relevant bases to zero. To compute the spline-based STRFs,  $w_{SPL}$ , the coefficients,  $b_{SPL}$ , were obtained as

$$b_{SPL} = (S^T X^T X S)^{-1} S^T X^T y \quad (\text{A.5})$$

with  $X$  as the stimulus design matrix,  $y$  as the neural response vector, and  $S$  as the spline matrix. The spatio-temporal RF was computed as

$$w_{SPL} = S b_{SPL} = S (S^T X^T X S)^{-1} S^T X^T y \quad (\text{A.6})$$

To generate the natural cubic spline matrices ( $S$ ) the package Patsy (0.5.1) is integrated into RFest.

To approximate  $w_{SPL}$  we used a generalized linear model (GLM) that predicts the instantaneous firing rate for one neuron using the movie as a predictor. We extended this "Stimulus only" model by integrating also running speed and pupil size as behavioral predictors of neuronal firing rate. To estimate the effect of cortico-thalamic (CT) feedback, an additional bimodal input was used comprising the optogenetic light stimulation that could be either on ( $o_t = 1$ ) or off ( $o_t = 0$ ). All inputs were parameterized with a set of spline basis and multiplied with an extra weight vector (also referred to as kernel):

$$f(s, o, r, e) = g(w_s^T s[t : t - \Delta t] + w_o^T o[t : t - \Delta t] + w_r^T r[t : t - \Delta t] + w_e^T e[t : t - \Delta t]) \quad (\text{A.7})$$

with  $s$ ,  $o$ ,  $r$ , and  $e$  denoting the additional model inputs of stimulus, optogenetics, running and eye, respectively, and  $[t : t - \Delta t]$  defining the temporal integration window (250 ms for stimulus and 800 ms for the predictors for modulatory inputs).

We estimated the kernels by gradient descent with respect to the cost function with L1 regularization. As standard procedure for time series data, we used 150 out of the 188 unique movie clips (80 %) for cross-validation to select optimal hyperparameters and reserved the remaining 38 clips (20 %) as a held-out test set. To select optimal hyperparameters, we used five-fold cross-validation grid search on the training data (120 training clips and 30 validation clips in each fold). Hyperparameters included the number of spline basis (between 10 and 19) in temporal dimension (for stimulus, pupil size, locomotion, and feedback input) and spatial dimension (only stimulus), as well as the strength of the L1 regularization (with weights varying from 5 to 15). The STRFs were initialized randomly and optimized using gradient decent for 2000 iterations. We stopped the fitting early when the training cost changed less than  $10^{-5}$  for 10 iterations. Finally, we selected the hyperparameters based on the mean performance on the validation set across folds. After optimal hyperparameter selection, models were retrained on the full training data (150 movie clips) and the final performance of the model was reported as the correlation coefficient between predicted and observed neural responses on the held out test data (38 movie clips).

We further fitted spline GLMs with responses to a sparse noise stimulus. We conducted a separate hyperparameter search for this stimulus, using a similar procedure as for the movie stimulus, involving cross-validation. For each neuron, we determined a unique set of optimal hyperparameters and kernels for the sparse noise stimulus and the movie stimulus.

## Session based permutation test

Following Harris (2020) [149], we performed a permutation test to evaluate if data from a different experimental session would lead to the same or worse model performance, which indicated the significance of the input-output correlations captured by the model. To achieve this, we provided the model with input data from the validation set of an unrelated session (for the model inputs ‘stimulus’, ‘running’, ‘pupil size’) or with synthetic inputs generated with the same statistics as the original (for the ‘CT feedback suppression’ input), one at a time. We repeated the process using all different data from all recording sessions. Subsequently, we compared for each input the actual model performance (Pearson’s  $r$ ) on the validation set across its folds ( $n=5$ ) against a distribution of model performances with that specific input permuted from the different recording sessions across their folds ( $n=5$  folds  $\times$  9 recording sessions). Inputs were considered significant if the actual performance differed from the permuted performance with  $p \leq 0.05$  using the non-paired Mann-Whitney-U test.

## Spatio-temporal RF component extraction

To separate spatial and temporal components of the 3D STRFs, we performed singular value decomposition (SVD) on the norm of the stimulus weight vector  $w$ . The temporal RF is extracted as the first left-singular vector of  $U$ , i.e. temporal dimension with the highest variance, and the spatial RF as the first right-singular vector of  $V$ , reshaped into the height- and width-dimensions of the input vector  $w$ . The extremes of the reshaped spatial RF vector are then used to quantify RF position and RF area. The extracted temporal RF components were normalized and multiplied with the RF center value before computing the slope (-150 ms to peak).

## Spatial RF contour estimation

Model spatial RFs were estimated by extracting the 2D spatial RF component from the model weights (see subsection ‘Spatio-temporal RF component extraction’), and then drawing a contour line around the largest absolute peak (assumed to be the center of the spatial RF). The contour threshold gets gradually lowered until any further decrease would result in a second contour around the second largest extremum (background irregularities considered as noise). To avoid overly large RFs in very clean spatial components (without any major second extremum), the contour threshold had to be 2 standard deviations above or below the mean. To improve estimate accuracy, the spatial RF component was upsampled 16-fold via cubic spline interpolation.

## Spatial RF area

The spatial RF areas were estimated by using the spatial RF component and contour (see subsection ‘Spatial RF contour estimation’) and calculating the number of pixels of the spatial RF contour mask in relation to the total number of pixels in the image frame, which was then scaled by the stimulus extent to obtain the value in squared degrees of

visual angle. To improve estimate accuracy, the spatial RF component was upsampled 16-fold via cubic spline interpolation.

## Spatial RF center-surround

stRFs were collapsed across the azimuth axis, and accordingly defined as a function of elevation and time. Then, we identified the time point with peak activity. Subsequently, centre regions were defined according to the spatial width of the peak activity, and surround regions were defined as a ring encircling the centre and extending up to trice the diameter or  $9^\circ$ . We then summed pixel intensities within the centre region and the surround regions separately and calculated surround-to-centre ratio as:

$$\text{surround-to-center} = \frac{2 * \sum \text{centre}}{\sum \text{surround}_1 + \sum \text{surround}_2} \quad (\text{A.8})$$

## Functional cell types clustering

To categorize neurons into functional subtypes, we employed dimensionality reduction (PCA) on their PSTH responses to a full-field light intensity step stimulus. Subsequently, we conducted clustering using a Gaussian mixture model on the principal components, resulting in the identification of four primary groups (Sustained ON, Sustained OFF, Transient, and Mixed). Nevertheless, we observed substantial response diversity even within these main groups. To address this, we further subdivided neurons within each group through empirical selection of cluster numbers, aiming to optimize silhouette scores and reduce the standard error of the PSTH mean response within the subclasses. This two-step clustering approach allowed capturing finer distinctions of response patterns.

## Rate of change (RoC) to CT feedback suppression

To quantify the temporal dynamics of the effect of suppressing CT feedback with optogenetic pulses, we calculated rate of change (RoC) in the neurons' responses based on both data and model prediction. For each case, we first identified the time point  $t_{min}$  where the slope of neurons response flipped its sign (e.g. after light onset, firing rates decreased, negative slope, and then it either stays constant or fluctuates around its new value and the slope becomes zero or flips its sign to positive). Then, we calculated the normalised change in responses  $MI_{CT\ FB\ supp.}$ ; (see subsection 'Optogenetic feedback suppression'). Finally,  $RoC$  was defined as:

$$RoC = \frac{MI_{CT\ FB\ supp.}}{t_{min}} \quad (\text{A.9})$$

### A.2.13 Decoding analysis

We used a two-alternative forced choice (2AFC) decoder [308] to analyze if stimulus discriminability is easier with or without feedback. Given two spike trains  $\{s_A, s_B\}$  in response to two movie clips  $\{A, B\}$  we computed the model log-likelihood for the correct  $((s_A, A), (s_B, B))$  and the incorrect pairing  $((s_A, B), (s_B, A))$ . We used the 'Full model' to predict spike trains in the two feedback conditions, (1) with the feedback component intact and (2) with the feedback component suppressed. In both conditions, we used 100 randomly selected 50 ms movie clips from the test set and their corresponding responses.

Using all possible pairs, we computed the model log-likelihood for the correct and the incorrect pairing in both feedback conditions. The log probability for the correct pairing of response  $s_A$  and stimulus  $A$  is defined as follows:

$$\log p(s_A|A) = \sum_t s_{A,t} \log \lambda_t(A_t) - \lambda(A_t) \quad (\text{A.10})$$

where  $\lambda$  is the instantaneous firing rate at time  $t$  predicted by the model. Analogously, we computed  $\log p(s_B|B)$ ,  $\log p(s_A|B)$ , and  $\log p(s_B|A)$ . A correct choice was made if  $\log p(s_A|A) > \log p(s_A|B)$  or  $\log p(s_B|B) > \log p(s_B|A)$ , respectively. We used the percentage correct over all possible pairs to quantify decoding performance. Finally, we computed the ratio of percentage correct for the feedback suppressed condition and the control condition. A ratio above one indicates better decoding performance in the feedback suppressed condition and a ratio below one in the control condition.

## B | Supplementary material for Study II: Zouridis and Schmors et al. (2024)

### **B.1 Supplementary Figures**

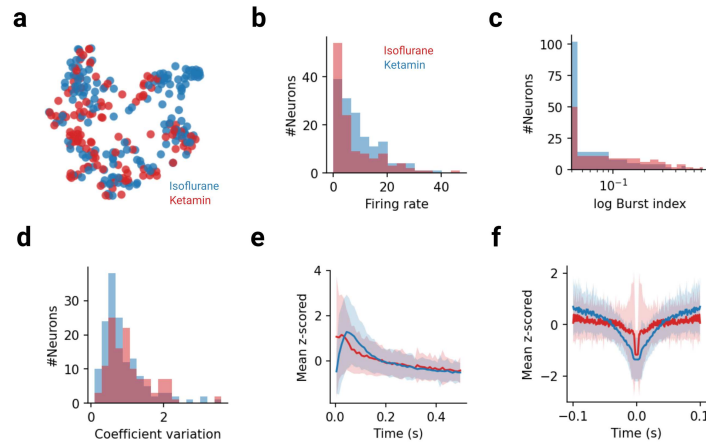


Figure B.1: **Comparison of LHB neurons recorded under ketamine or urethane anesthesia.** (a) t-SNE embedding of electrophysiological properties of lateral habenula (LHB) neurons recorded under ketamine (red;  $n_{\text{neurons}} = 149$ ) or urethane (blue;  $n_{\text{neurons}} = 121$ ) anesthesia. Each point represents a recording. (b) Mean firing rate distribution of neurons recorded under ketamine or urethane anesthesia. (c) Burst index distribution of neurons recorded under ketamine or urethane anesthesia. While NMDA antagonism has been shown to abolish bursting in LHB neurons, we observed Type 1 patterns under ketamine anesthesia. As expected, bursting was more prominent under urethane anesthesia. (d) Coefficient of variation of neurons recorded under ketamine or urethane anesthesia. (e) Mean interspike interval distribution of neurons computed for neurons from ketamine and urethane condition. Shaded area indicates  $\pm 1$  SD. (f) Same as (e) for mean autocorrelograms.

B. Supplementary material for Study II:  
Zouridis and Schmors et al. (2024)

---

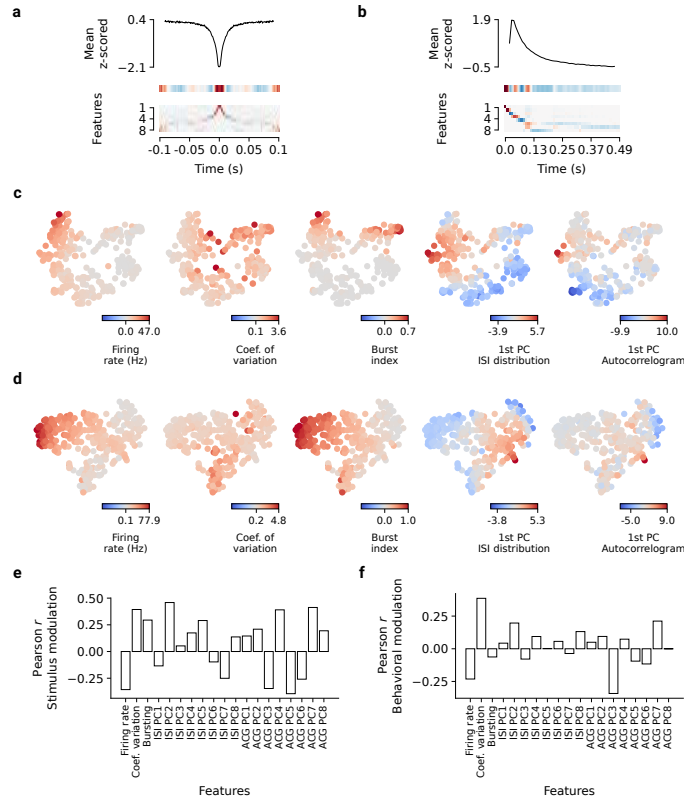


Figure B.2: **Distributions of features and their correlations across LHB neurons in anesthetized and awake conditions.** **(a)** ACG feature extraction using sparse Principal Components Analysis (sPCA). *Top*: mean z-scored ACG for all neurons (black;  $n_{\text{neurons}} = 565$ ). *Bottom*: all 8 sPCA components. **(b)** Same as in (a) but for interspike-interval. **(c)** t-SNE embedding of electrophysiological properties of LHB neurons recorded under anesthesia. Neurons are color coded by the respective feature, from left to right: firing rate, coefficient of variation, burst index, 1st PC ISI distribution, 1st PC Autocorrelogram. **(d)** Same as in (c) but for neurons recorded under anesthesia. **(e)** Pearson  $r$  computed for the modulation by FS stimulus with clustering features. See (a) and (b) for interpretation of PC features. **(f)** Same as in (e) but for modulation by spontaneous behavior.

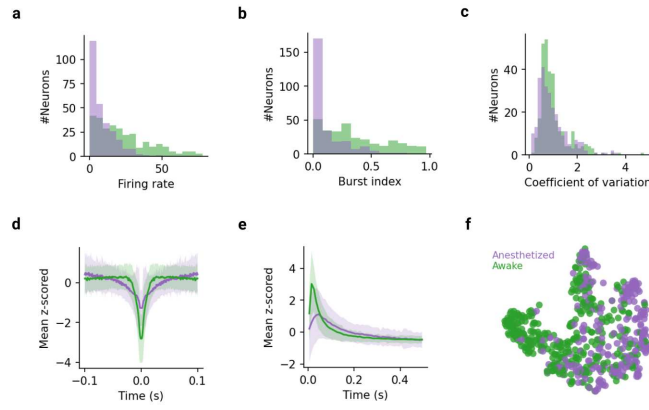


Figure B.3: **Comparing LHB neurons recorded under anesthetized condition and awake condition.** (a) Firing rates of all LHB neurons in the anesthetized data set (purple;  $n_{\text{neurons}} = 270$ ) in the awake data set (green;  $n_{\text{neurons}} = 295$ ). (b) Burst index comparison of neurons from anesthetized and awake condition. (c) Comparison of coefficient of variation computed for neurons from anesthetized and awake condition. (d) Comparison of mean interspike interval distribution computed for neurons from anesthetized and awake condition. Shaded area shows  $\pm 1$  SD. (e) Comparison of mean autocorrelograms (ACG) computed for neurons from anesthetized and awake condition. Shaded area shows  $\pm 1$  SD. (f) t-SNE embedding of electrophysiological properties of lateral habenula (LHB) neurons recorded under anesthesia (purple;  $n_{\text{neurons}} = 270$ ) or awake state (green;  $n_{\text{neurons}} = 295$ ). Each point represents a neuron.

## B.2 Materials and methods

### B.2.1 Experimental animals.

Experiments were performed on adult male C57BL/6J mice (>8 weeks old, 20–30 g body weight; Charles River, Sulzfeld, Germany). The mice were housed under a 12-hour light cycle and were provided with unrestricted access to food and water. All experimental procedures complied with the German Animal Welfare Act (TierSchG) and Animal Welfare Laboratory Animal Ordinance (TierSchVersV). The study received approval from the state authority (Regierungspräsidium, Tübingen, Baden-Württemberg, Germany).

### B.2.2 In vivo juxtacellular recording and labeling in anesthetized mice.

In vivo juxtacellular recordings of LHb neurons in mice have been compiled from previous studies; the procedures and datasets are described in detail in [84]. Briefly, recordings include juxtacellular recordings of spontaneous activity of LHb neurons in mice anesthetized with either a ketamine-xylazine mixture or urethane. In a subset of mice, foot shock (FS) stimulation was applied to the contralateral hind limb. A comparative analysis indicated that neuronal firing properties were comparable between ketamine-xylazine and urethane anesthetized mice (Figure B.1), allowing us to combine recordings obtained under different anesthetics into a single dataset.

### B.2.3 In vivo juxtacellular recordings and orofacial videography in awake head-fixed mice.

**Surgical procedures.** Surgery, habituation and recordings in awake head-fixed mice were essentially performed as previously described [28, 42]. Briefly, mice were anesthetized with a ketamine/xylazine mixture (80-100 and 10 mg/kg, respectively, i.p.), fixed on a stereotaxic frame (Narishige, Tokyo, Japan), and a custom-made head-post was attached to the skull using a light-curing adhesive (Optibond Universal, Kerr) and dental acrylic (Paladur, Heraeus Kulzer, Hanau, Germany). For targeting the LHb, a craniotomy was performed at the stereotactic coordinates: 1.5 mm posterior and 0.5 mm lateral to bregma. In between recording sessions, the craniotomy site was covered with a silicon sealant (Kwik-Cast, WPI).

**In vivo juxtacellular recording and labeling.** A neurobiotin-filled [1.5–2% Neurobiotin, SP-1120, Vector Laboratories in ‘intracellular-like’ solution (in mM): 135 K-gluconate, 10 HEPES, 10 Na<sub>2</sub>-phosphocreatine, 4 KCl, 4 MgATP, and 0.3 Na<sub>3</sub>GTP, osmolarity: 280–310 mOsm] glass pipette electrode [(Cat#1403547, Hilgenberg, Malsfeld, Germany; pulled to a 7-9 MΩ resistance with a micropipette puller (Cat#P-1000, Sutter Instrument, Novato, CA, USA)] was lowered into the LHb (coordinates: 2.3-2.9 mm ventral to the brain surface). Juxtacellular voltage signals were acquired using an ELC-03XS amplifier (NPI Electronic, Tamm, Germany) and digitized at 25 kHz (Power1401-3, Spike2 v8.02; CED, Cambridge, UK). Upon establishing juxtacellular configuration with a neuron, spontaneous spiking activity was monitored. When feasible, once the recording was completed, neurons were labeled using current pulses [311].

**Histological verification of the recording location.** To confirm the recording location in the LHb, animals were overdosed with pentobarbital (300 mg/kg, i.p.), transcardially perfused (0.1M phosphate buffer followed by a 4% paraformaldehyde 0.1 phosphate buffer solution), and dissected. The brain tissue was processed essentially as previously described [464]. Briefly, serial tissue sections (70  $\mu$ m thickness; coronal sectioning plane) were obtained using a vibrating blade microtome (VT1200, Leica, Germany), stained with a fluorescent streptavidin conjugate (1:1000, Cat#S11225, Thermo Fisher Scientific) and DAPI (1:1000, Cat#D1306, Thermo Fisher Scientific), and examined using epifluorescence microscopy (Axio Imager, Carl Zeiss, Germany). Recordings from a session were included in analysis only if labeled cells or electrode tracks were within the LHb.

**Orofacial videography.** The mouse's left whisker pad was illuminated with infrared light (M850L3, Thorlabs), video was recorded using a camera (DMK33UX265, TheImagingSource) externally triggered, and trigger pulses were co-registered with electrophysiology data (see B.2.3).

#### B.2.4 Quantification and statistical analysis.

Data analysis was performed using custom-written code that applies general tools such as Numpy, sklearn, matplotlib, seaborn, and pandas. All statistical analyses were conducted using Python scikit-learn library. To compare more than two groups we used ANOVA with  $\alpha = 0.05$ . In total we included  $n_{\text{neurons}} = 88$  from Congiu et al. (2023) [84],  $n_{\text{neurons}} = 33$  from Lecca et al. (2017) [215], and  $n_{\text{neurons}} = 149$  from Congiu et al. (2023) [84], which yielded to  $n_{\text{neurons}} = 270$  from anesthetized animals. For recordings from awake mice we had  $n_{\text{neurons}} = 295$ .

**Electrophysiology data analysis.** The analysis of juxtacellular electrophysiological data was essentially performed as previously described (e.g., [28, 42, 464]). Specifically, spikes were isolated from continuous voltage traces using a threshold-based approach and visual inspection, as previously described [464]. For each neuron, we computed the mean firing rate (FR), the interspike interval (ISI) distribution (range 0–0.5 s), the coefficient of variation (CV; calculated as the ratio of the standard deviation of the ISI over the mean ISI [373]), the burst index (calculated as the percentage of ISIs shorter than 25 ms), and the autocorrelogram (ACG; range  $\pm 0.1$  s) based on spontaneous activity. Temporal features of the ISI distributions and ACGs were extracted with sparse principal component analysis (sPCA; [463]) using the first 8 principal components (PCs; Figure B.2a,b).

**Clustering.** We fit each data set with a Mixture of Gaussians model [40] using the following eleven features: mean FR, CV, burst index, first 8 PCs of the ISI distribution, and first 8 PCs of the ACG (Figure B.2a,b). We applied the commonly used Bayesian Information Criterion [126] to determine the optimal number of clusters:  $BIC = -2 \log[L] + M \log[N]$ , with  $L$  as the likelihood function for the model, and  $M$  as the number of parameters in the model. The BIC naturally implements Occam's razor by favoring simpler models (fewer clusters) unless there is strong evidence in the data for more complex models (more clusters). To avoid local minima, we initialized the GMM algorithm 100 times per candidate cluster number (error bars in Figure B.2c) and selected the number of clusters with the smallest mean BIC which resulted in four clusters for both data sets. To match clusters from the anesthetized preparation (A) with clusters from the awake preparation (B), we used the cosine similarity  $S_C$  [371] to measure the similarity between cluster centroids:  $S_C(\text{cluster}_A, \text{cluster}_B) = \frac{A \cdot B}{\|A\| \|B\|}$ . We applied the Hun-

garian algorithm [206, 276] to find the optimal matching between clusters from the two datasets. The algorithm optimally matches clusters by minimizing the total dissimilarity  $\min_{\pi \in \Pi} \sum_{i=1}^n (1 - S_C)$ , where  $\pi$  is a permutation of cluster indices and  $\Pi$  is the set of all possible combinations.

**Analysis of foot shock responses.** To assess responses of neurons to foot shock stimuli, we constructed peri-stimulus time histograms (−0.2 to +1.3 s with respect to stimulus onset; bin width, 20 ms; critical firing rate > 1 Hz; **Figure 3.7a**). To determine the response latency to foot shock, we calculated the median latency of the first spikes occurring within 100 ms after stimulus onset. To quantify the strength of foot shock modulation, we computed a foot shock modulation index (FSMI) [28], defined as:

$$\text{FSMI} = \frac{\text{FR}_{\text{stimulus}} - \text{FR}_{\text{baseline}}}{\text{FR}_{\text{stimulus}} + \text{FR}_{\text{baseline}}}$$

where  $\text{FR}_{\text{baseline}}$  represents the mean firing rate during 1 s preceding the stimulus onset,  $\text{FR}_{\text{stimulus}}$  represents the mean firing rate during stimulus delivery.

**Videography data analysis.** To assess the behavioral state, we used gross facial motion as an index of arousal. Face motion energy and pupil diameter were extracted using Facemap [385] as detailed in [42]. To partition behavior into states, we defined active immobility epochs as those for which whisker pad motion energy exceeded the 80th percentile; the rest were defined as inactive. For every neuron, we tested if the firing rate modulation was significant using an unpaired Wilcoxon signed-rank test with a significant level of 0.05. Neurons were split into neurons that show a significant increase or decrease in firing rate and neurons that showed no significant change in firing rate depending on facial motion. To quantify the strength of modulation, we computed a facial motion modulation index (FMMI) as:

$$\text{FMMI} = \frac{\text{FR}_{\text{whisking}} - \text{FR}_{\text{control}}}{\text{FR}_{\text{whisking}} + \text{FR}_{\text{control}}}$$

where  $\text{FR}_{\text{control}}$  represents the mean firing rate during active periods,  $\text{FR}_{\text{whisking}}$  represents the mean firing rate during inactive periods.

## C | Supplementary material for Study III: Congiu et al. (2023)

### **C.1 Supplementary Figures**

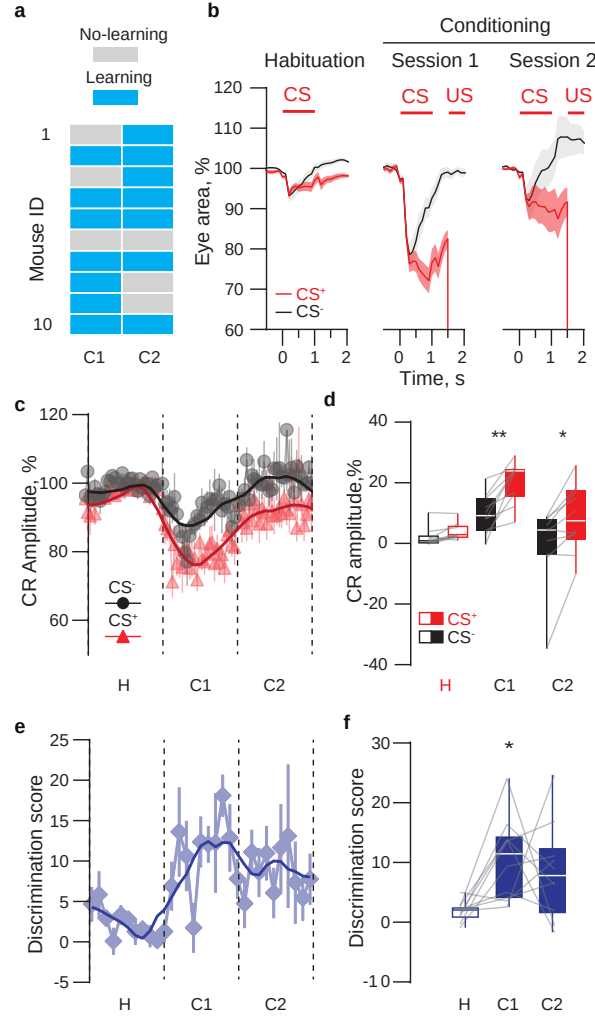


Figure C.1: **Features of the Pavlovian discrimination task.** (a) Categorical heatmap depicting mice performance during conditioning session 1 and 2. (b) Time course of the average eye area changes across CS+ and CS- trials during habituation (left) and conditioning (session 1: center, session 2: right) ( $n_{\text{mice}} = 10$ ). (c) Time course of the normalized eye closure for CS+ and CS- trials (bin = 3) during habituation and conditioning sessions. (d) Boxplots of the conditioned response amplitude for CS+ and CS- across habituation and conditioning sessions ( $n_{\text{mice}} = 10$ ; Two Way ANOVA RM and Sidak's multiple comparison test, CS+ vs CS-  $F(1,9)=87.66$ ; H:  $p=0.99$ ; C1:  $**p=0.022$ ; C2:  $*p=0.018$ ). (e) Averaged time course of the discrimination score (bin = 3) during habituation and conditioning sessions ( $n_{\text{mice}} = 10$ ). (f) Boxplots of the discrimination score habituation and conditioning sessions ( $n_{\text{mice}} = 10$ ; Two Way ANOVA RM;  $F(1.378, 12.38) = 4.31$ ,  $*p=0.049$ ).

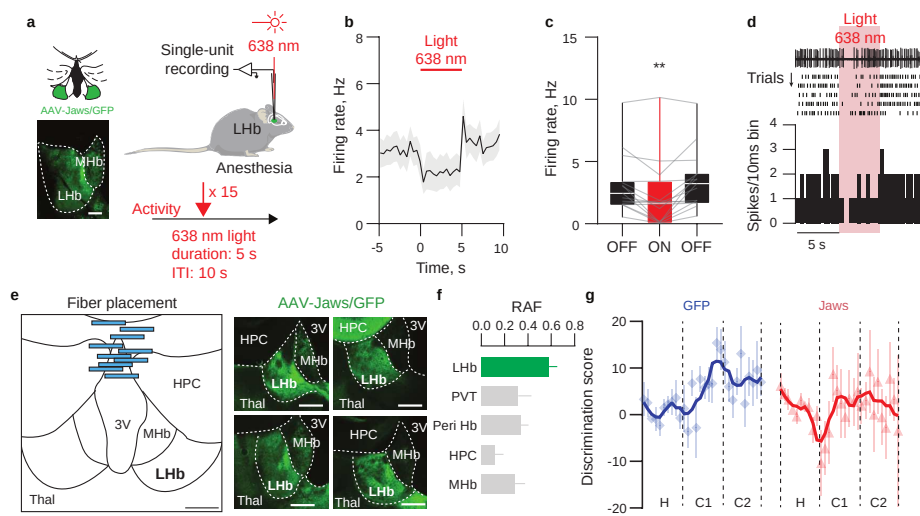


Figure C.2: **Optical inhibition of LHb during the Pavlovian discrimination task.** (a) Schematic of the experimental setting for the infection of the AAV-Jaws in the LHb and example histological coronal section of the expression of Jaws in the LHb (left; scale bar: 200  $\mu$ m) and schematic of the experimental protocol for the electrophysiological recordings of LHb neurons in anesthetized mice. (b) Time course of the average firing rate of the recorded LHb neurons during light delivery ( $n_{\text{mice}} = 3$ ,  $n_{\text{neurons}} = 15$ ). (c) Boxplot of the average firing rate before, during and after red light presentation ( $n_{\text{mice}} = 3$ ,  $n_{\text{neurons}} = 15$ ; Two Way ANOVA RM;  $F(1.06, 14.78) = 12.46$ ,  $**p=0.003$ ). (d) Example trace, raster plot and peristimulus time histogram of an example neuron inhibited by light presentation. (e) Fiber placement for all Jaws mice (left) and additional example pictures of Jaws expression in the experimental mice (right; scale bar: 200  $\mu$ m). (f) Bar-graph representing the average relative arbitrary fluorescence measured in the LHb and neighboring regions ( $n_{\text{mice}} = 10$ , one-way ANOVA RM,  $F(2.520, 22.68) = 12.81$ ,  $*p<0.0001$ ) (g) Averaged time course of the discrimination score (bin = 3) during habituation and conditioning sessions for GFP (left) and Jaws (right) expressing mice ( $n_{\text{mice}} \text{ GFP} = 7$ ;  $n_{\text{mice}} \text{ Jaws} = 10$ ; habituation GFP vs Jaws: Kolmogorov-Smirnov test,  $D = 0.101$ ,  $p=0.168$ ; conditioning GFP vs Jaws: Kolmogorov-Smirnov test,  $D = 0.183$ ,  $***p<0.001$ ).

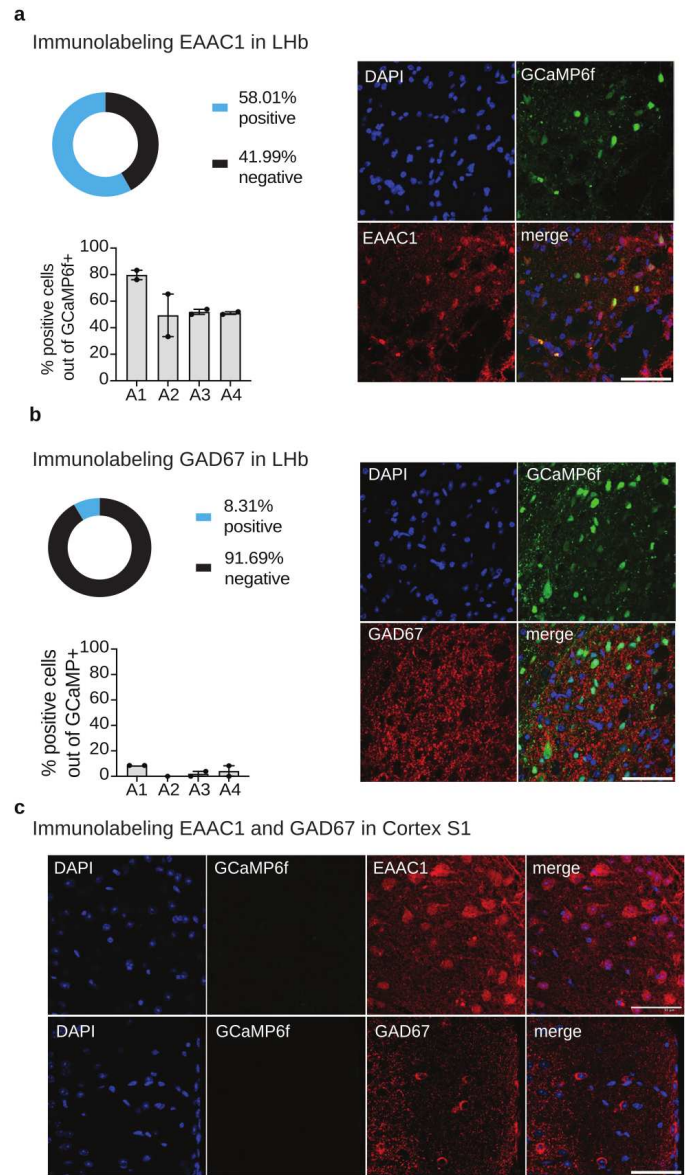


Figure C.3: **Immunolabeling of glutamatergic and GABAergic neurons in the LHb.** (a) Representative confocal z-stack projections of LHb from mice infected with AAV-GCaMP6f, and relative quantification of GCaMP6f fluorescent cells co-localizing with EAAC1 (shown as % of positive cells out of GCaMP6f-positive total cells). (b) Same as A but for GAD67 immunoreactivity. (c) Representative confocal z-stack projections of EAAC1 and GAD67 immunoreactivity in the somatosensory cortex of the same mice ( $n_{\text{mice}}=4$ ). Scale bar= $30\text{ }\mu\text{m}$ .

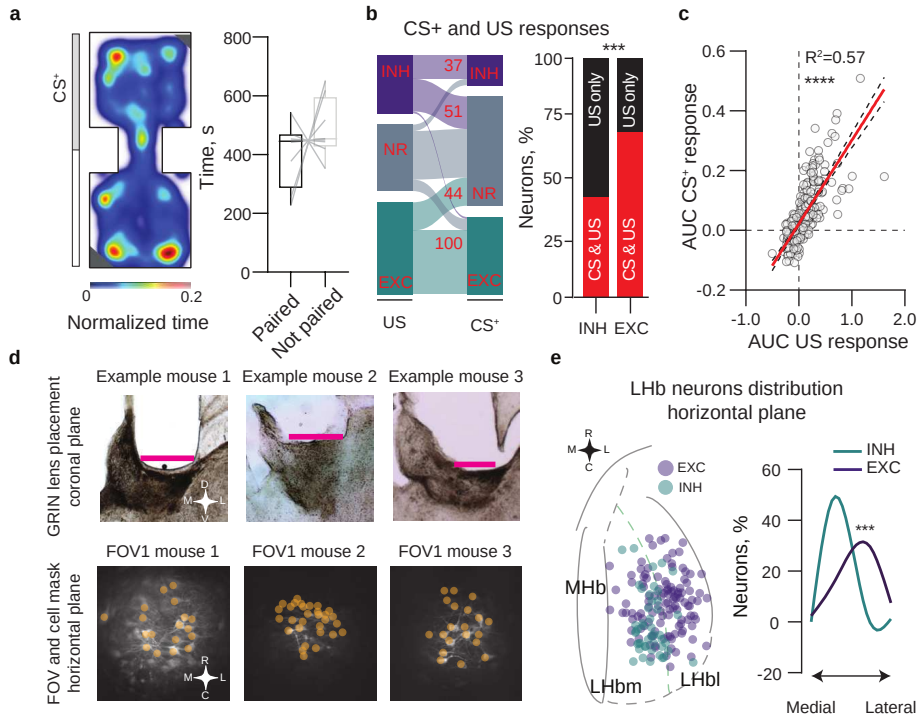


**Figure C.4: Functional and anatomical signatures of LHb neurons during conditioning.** (a) Example heatmap for a single mouse depicting time spent in the compartment paired with the CS+ auditory tone vs the not paired one. Box-plots and individual values reporting time in compartments ( $n_{\text{mice}} = 8$ ; paired t-test;  $t_7 = 1.171$ ,  $p = 0.280$ , paired t test). (b) Sankey-diagram representing neuronal responses to the US and CS+ during conditioning (left) and bar graph representing the quantification of the percentage of neurons that are responding to both CS+ and US or only US (right;  $X^2 = 14.76$ , \*\*\* $p < 0.0001$ ). (c) Scatter plot of the single neuronal response for CS+ and US ( $n_{\text{mice}} = 12$ ,  $n_{\text{neurons}} = 339$ ; Pearson correlation,  $R^2 = 0.566$ ,  $F(1,337) = 440.2$ , \*\*\* $p < 0.0001$ ). (d) Histological example sections of GRIN lens placement above the LHb (top) and respective field of view (FOV) average projection (bottom). (e) Anatomical reconstruction of excited and inhibited neuron location in horizontal plane and relative quantification of the medio-lateral spatial distribution ( $\chi^2 = 66.56$ , \*\*\* $p < 0.0001$ ).

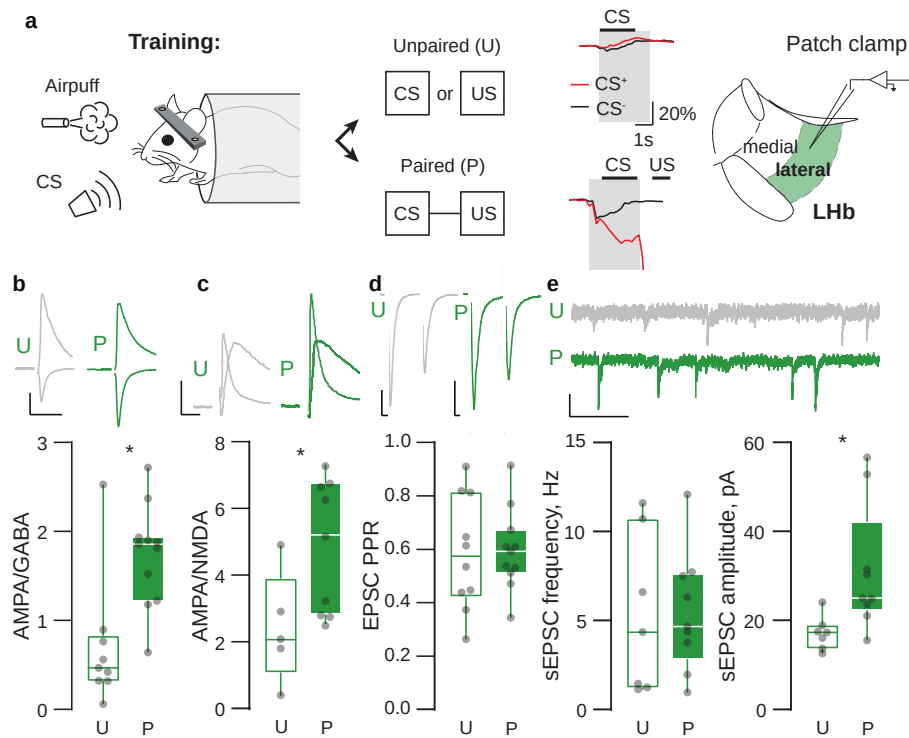


Figure C.5: **Potentiation of synaptic excitation in the LHb after learning.** (a) Schematic of the experimental setting for the behavioral training (left), training protocol for the two experimental groups and example of behavioral performance (center), and schematic representation of the location of the electrophysiological patch clamp recordings (right). (b) Example traces (top; scale bars: 50pA, 50 ms) and boxplots (bottom) of the AMPA/GABA ratio ( $n_{\text{mice}}$ : U = 2, P = 3,  $n_{\text{neurons}}$ : U = 9, P = 11; unpaired t-test,  $t_{18} = 3.53$ ,  $*p=0.0024$ ). (c) Example traces (top; scale bars: 50pA, 10ms) and boxplots (bottom) of the AMPA/NMDA ratio ( $n_{\text{mice}}$ : U = 3 P = 4,  $n_{\text{neurons}}$ : U = 5, P = 9; unpaired t-test,  $t_{12} = 2.323$ ,  $*p=0.039$ ). (d) Example traces (top; scale bars: 50 pA, 10 ms) and boxplots of the EPSC PPR (bottom;  $n_{\text{mice}}$ : U = 2, P = 2,  $n_{\text{neurons}}$ : U = 10, P = 11; unpaired t-test,  $t_{18} = 0.14$ ,  $p=0.8941$ ). (e) Example traces (top; scale bars: 250 ms, 20 pA), boxplots of the spontaneous EPSC frequency (bottom left;  $n_{\text{mice}}$ : U = 2, P = 3;  $n_{\text{neurons}}$ : U = 7, P = 9; unpaired t-test,  $t_{14} = 0.09$ ,  $p=0.92$ ) and boxplots of the spontaneous EPSC amplitude (bottom right;  $n_{\text{mice}}$ : U = 2, P = 3;  $n_{\text{neurons}}$ : U = 7, P = 9; unpaired t-test,  $t_{14} = 2.55$ ,  $*p=0.023$ ).

C. Supplementary material for Study III:  
Congiu et al. (2023)

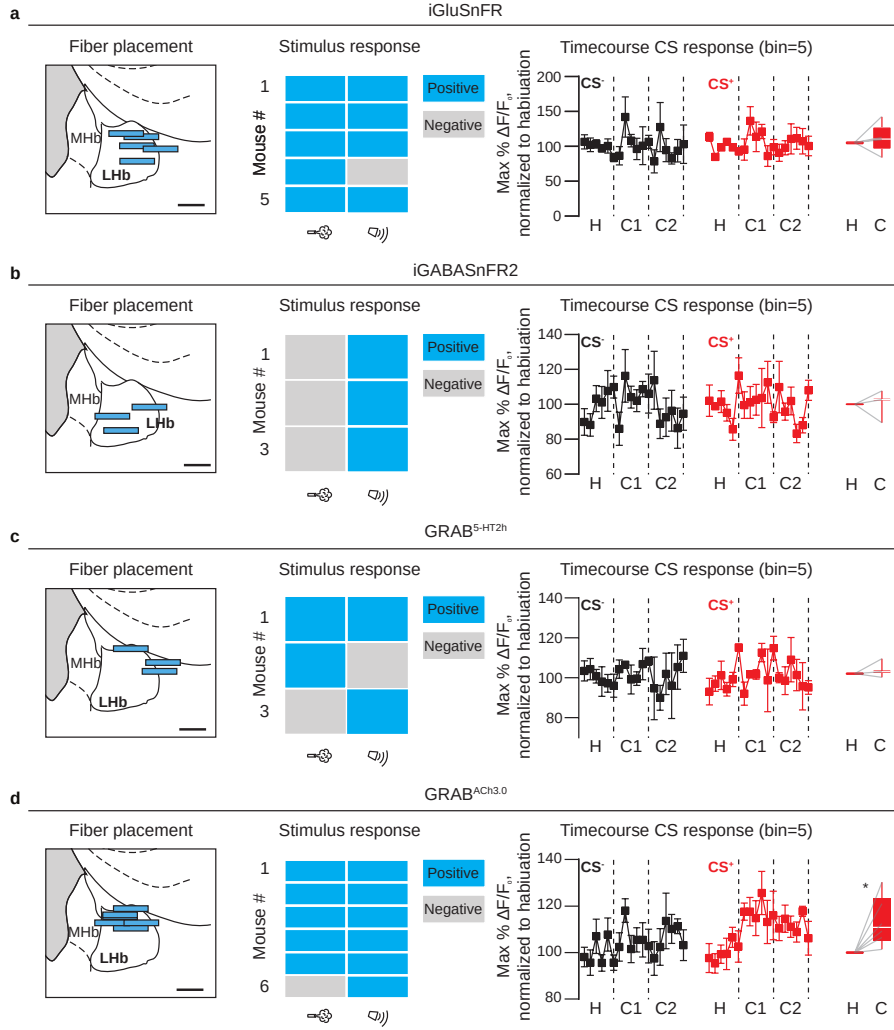


Figure C.6: **Analysis of neurotransmitter dynamics in the LHb.** (a-d) Left, fiber placement (left, scale bar: 200  $\mu$ m); centre left, binary heatmap representing single mouse significant response to unpredicted airpuff and cue presentation during the habituation session (responses were considered significant when the average photometric responses was higher or lower than baseline for at least 150 ms); centre right, Time course of the maximum fluorescence (normalized to habituation) detected during CS+ and CS- presentation across training; right, boxplot depicting the maximum fluorescence detected during cue presentation normalized to the habituation session (Paired t-test. iGluSnFR:  $n_{\text{mice}} = 5$ ;  $t_4 = 0.721$ ,  $p=0.511$ . iGABASnFR2:  $n_{\text{mice}} = 3$ ;  $t_2 = 0.010$ ,  $p=0.993$ . GRAB5-HT2h:  $n_{\text{mice}} = 3$ ;  $t_2 = 0.81$ ,  $p=0.503$ . GRAB-ACh3.0:  $n_{\text{mice}} = 6$ ;  $t_5 = 3.14$ ,  $*p=0.026$ ).

## C.2 Materials and methods

### C.2.1 Experimental subjects

C57BL/6 J wild-type male mice 7–25 weeks old were used for this study (Janvier lab, France; Charles River, Sulzfeld, Germany). Mice were housed in three to five per cage with water and food ad libitum on a 12:12 h light cycle (lights on at 7 a.m.) in individually ventilated cages (ICV, Innovive, France). All experimental procedures were approved by local authorities and were performed according to guidelines of the respective local ethics committee: the canton of Vaud Cantonal Veterinary Office Committee for Animal Experimentation (Switzerland; License VD3171), in compliance with the Swiss National Institutional Guidelines on animal experimentation; the Regierungspräsidium (Tübingen, Baden-Württemberg, Germany; License CIN07/19G and CIN03/20G) in compliance with the German Animal Welfare Act (TierSchG) and the Animal Welfare Laboratory Animal Ordinance (TierSchVersV).

### C.2.2 Viruses

rAAV-DJ/8/2-hSyn1-eGFP-WPRE (titer:  $9.4 \times 10^{12}$  vg/ml), rAAV-DJ/8/2-hSyn1-GCaMp6f-WPRE (titer:  $4.5 \times 10^{12}$  vg/ml), ssAAV-5/2-hSyn1-iGluSn-FR(A184S)-WPRE (titer:  $5.3 \times 10^{12}$  vg/ml) were purchased from the UZH Vector Facility (Zurich, Switzerland). AAV9-hSyn-5HT3.5 (titer:  $1 \times 10^{13}$  vg/ml), AAV9-hSyn-ACh3.0 (titer:  $1 \times 10^{13}$  vg/ml) were purchased from BrainVTA (Wuhan, China). pGP-AAV5-syn-iGABASnFR2-WPRE (titer:  $2.67 \times 10^{13}$  vg/ml) was provided by the GENIE Project at HHMI Janelia Research Campus (<https://doi.org/10.25378/janelia.19709311.v3>). rAAV8-hSyn1-Jaws-GFP (titer:  $1.3 \times 10^{13}$  vg/ml) was purchased from Addgene.

### C.2.3 Stereotaxic surgeries for viral injections

Mice were anesthetized with ketamine (150 mg/kg)/xylazine (10 mg/kg) (Cantonal University Hospital, Lausanne, Switzerland). The surgery was performed using the ocular protector Viscotear to prevent eye damage, a heating pad to keep a stable body temperature, and local anesthesia (subcutaneous injection) with a mix of lidocaine (6 mg/kg) and bupivacaine (2.5 mg/kg). We unilaterally or bilaterally (when required for the experiment, i.e. optogenetics in vivo) injected in the LHb (−1.4 mm AP, 0.45 mmML, 3.1 mm DV) 150 – 250 nl of virus using a glass pipette on a stereotactic frame (Kopf, France). All injections were performed at a rate of approximately 100 – 150 nl/min. The injection pipette was withdrawn from the brain 10 min after the infusion. Animals were allowed to recover for a minimum of two weeks before fiber or GRIN lens implantation.

### C.2.4 Chronic implants

For fiber photometry experiments, a single fiber probe (200  $\mu$ m, Chi Square Bioimaging) was placed and fixed (C and B Metabond, Parkell) 100  $\mu$ m above the injection site in the LHb. For optogenetic manipulation a single fiber (200  $\mu$ m, Thorlabs) was placed at the following coordinates (AP: −1.4 mm, L:  $\pm$  0.1 mm, V: −2.2 mm). Surgery was performed under isoflurane anesthesia (induction: 4%, maintenance: 1.8 – 2%, Univentor). For endoscope experiments, mice were anesthetized (as described above) and implanted with a GRIN lens (6.1 mm length, 0.5 mm diameter; Inscopix, #100-000588). The lens was placed  $\approx$ 50–200  $\mu$ m above the injection site using the following coordinates from

bregma (−1.40 mm AP, 0.45 mmML, −2.85 to −2.9 mm DV; lowered at a speed of 1  $\mu\text{m/s}$ ). A stainless steel headbar was implanted on the skull. To do so, the skull was scraped clean and covered with a layer of Cyanoacrylate glue (Vetbond, 3 M). The headbar was lowered to touch the skull over lambda, then secured to the skull with a layer of dental adhesive (C and B Metabond, Parkell), followed by dental cement (Jetkit, Lang). For pain management, paracetamol (500 mg/250 ml; 200-300 mg/kg/day) was added to the drinking water after the surgery. Proper viral expression and fiber/GRIN lens placement in brain areas of interest were confirmed post hoc using histology for all experiments.

### C.2.5 Histology and immunohistochemistry

For histology, mice were terminally anesthetized with ketamine and xylazine or pentobarbital and perfused transcardially with paraformaldehyde (PFA) 4% in 0.1M phosphate-buffered saline (PBS). Brains were collected and left overnight in 4% PFA at 4 °C until slicing. Consecutive coronal slices (60–100 microns) were sectioned using a vibratome (Leica VT1200S). We took images with an epi-fluorescent microscope (Zeiss) to confirm efficient targeting of the LHb, and we discarded mice whenever this was not achieved (fiber optic, GRIN lens misplacement, or not targeted viral expression). Juxtacellularly labeled neurons (as in [Figure 3.8b,c](#)) were visualized according to previously published procedures [28]. Specifically, brains were sliced on a vibratome (VT1200S; Leica) to obtain 50–70  $\mu\text{m}$  thick parasagittal or coronal sections. Brain slices were processed with streptavidin conjugates (streptavidin-546 or -488, Invitrogen). Fluorescence images were acquired by epifluorescence microscopy (Axio imager; Zeiss) and confocal microscopy (LSM 900; Carl Zeiss, Oberkochen, Germany). To calculate a relative arbitrary fluorescence (RAF) of Jaws expression in LHb and neighboring regions, green signal intensity was normalized against a region containing no green fluorescence (background) using the formula:  $(\text{signal} - \text{background}) / (\text{signal} + \text{background})$ . For immunohistochemical analysis, brain sections were permeabilized at room temperature (RT) in 0.5% Triton X-100 (Sigma), followed by 1 h RT blocking in 2% BSA 0.5% Triton X-100 and overnight incubation with primary antibodies (EAAC1, AB1520 Millipore, 1:100; GAD67, MAB#5406 Clone 1G10.2, 1:1000) at 4 °C. Upon washing, sections were incubated for 2 h RT with Alexa-fluorophore-conjugated secondary antibodies (Invitro-gen). Confocal microscopy was performed with a Stellaris 5 (Leica) Laser Scanning System and images were processed and analyzed by FIJI/ImageJ Software. Confocal z-stack projections of confocal micrographs for GCaMP6f-GFP signal and for EAAC1 or GAD67 immunofluorescence were thresholded and used to manually outline positive cells. Cell masks produced for GCaMP6f and EAAC1 or GAD67 were subsequently multiplied, to generate the counts of double positive cells.

### C.2.6 Pavlovian conditioning

For Pavlovian conditioning, mice were headfixed in a 3 cm-wide acrylic cylinder. The setting was adapted from eye-blink conditioning methodological preparations as follows. A Basler camera (a2A1920-160umPRO or acA1300-60gc) was equipped with a macro-objective (either Fujinon HF9HA-1S or computar M3Z1228C-MP), and infrared LEDs were used to illuminate mice's face. For behavioral assessment during two-photon calcium imaging, the camera was equipped with a NIR short-pass filter with a 900 nm cutoff (FES0900, Thorlabs) to filter out excitation light. The airpuff was delivered with pressure from a Picospritzer III (Parker) with a 23G needle directed to the right eye (1 cm distance). A speaker to deliver acoustic tones was placed nearby the mice and was controlled by an MP3 trigger WIG-13720 (Sparkfun). The behavioral apparatus was controlled via Arduino custom code. Behavioral task events were transferred onto Excel with

PLX-DAQ interface. Facial videography was recorded with a frame rate of 10 Hz. Before behavior started, mice were handled (1-2 days) and acquainted to head fixation (1-2 days). Behaviour started with 2 sessions of habituation where mice were presented with 30 CS+ (16 kHz trilling sound, 75 dB, duration: 1 s), 30 CS- (10 kHz pure tone, 75 dB, duration: 1 s) and 5 unpredicted US (airpuff, pressure: 4-5 psi, duration: 0.5 s), delivered in random order with an inter-trial interval between 30 and 45 seconds. During conditioning sessions (2-3 days), mice received a total of 30 CS- trials and 30 CS+ trials where US was followed after a delay of 0.5 s from the end of CS. For two-photon recordings, mice were submitted to extra sessions of habituation and conditioning (1 to 5) to maximize data collection. For in vitro recordings, the paired group (P) was trained as described while the unpaired group (U) received the same number of stimuli with the same characteristics, but CS+ and US were never contingent (30 CS+, 30 CS- and 30 unpredicted US trials). For Jaws- mediated optogenetic silencing, 638 nm light (3 mW) was delivered by a laser (Matchbox Integrated Optics) to the LHb through a fiber cable (M83L1, Thorlabs) connected to a fiber optic implanted in the head of the mouse (CMFLC12L05, Thorlabs). Light was delivered during CS+ trials for 5 s starting from cue presentation. Behavioral assessment was recorded for a total of 15 s each trial with a sample rate of 10 fps starting 5 s before CS or US delivery. Data were analyzed using a custom-made Python script by thresholding the images to get only the eye pixel values. These values were normalized to 100 percent on a trial-by-trial basis in reference to the average value during the 5 s before cue presentation. From these data, the conditioned response (CR) amplitude was computed by averaging the values between the start of cue delivery (time: 0 s) and the start of airpuff presentation (time: 1.5 s). The discrimination score was obtained by subtracting the average value of the CR for CS+ trials to the one of CS- trials.

## C.2.7 Two-photon microscopy

Two-photon microscopy experiments were carried out to visualize LHb neuronal Calcium dynamics in vivo. Recordings and Pavlovian conditioning started when the field of view (FOV) was clear and stable for at least 3 weeks, typically 1-3 months after GRIN lens implantation. The acquisition was performed with a two-photon Ultima Investigator microscope (Bruker) combined with a Chamaleon-Ultra Tunable Laser (Coherent) and equipped with an Olympus 20× air objective (LCPLN20XIR). The two-photon microscope is equipped with a hybrid scanning core set with galvanometers and fast resonant scanners (up to 30 Hz frame-rate acquisition), multi-alkali PMT and GaAsP-PMT photodetectors with adjustable voltage, gain, and offset features, a single green/red NDD filter cube. Images were acquired at 30 Hz at a resolution of  $512 \times 512$  pixels and we performed an online averaging of 6 times to get an actual frame rate of 5 Hz. The average power of the beam measured at the front of the objective was between 80 and 125 mW. At each trial, recordings started 5 s before cue delivery, and a 15s-long video was collected. Images were collected using a computer equipped with Prairie view (Bruker). Prior experiment started, 1 up to 3 FOV per mouse was chosen to maximize the number of neurons recorded each session by adjusting the imaging plane (z-axis), and each FOV was spaced at least 60  $\mu\text{m}$  from one another to prevent visualization of the same cells across multiple FOVs. For imaging sessions with multiple fields of view, the conditioning session was split, allowing an equal number of trials per FOV. Sessions in which mice did not display any behavioral learning phenotype, were excluded from the analysis. After data collection, videos were motion-corrected with a planar hidden Markov model (SIMA v1.3.2), and regions of interest (ROIs) were manually drawn around each cell using the standard deviation projection of the motion-corrected video using ImageJ.19. Neurons that were not tracked for at least one habituation and one conditioning session were not kept in the analysis. Next, fluorescence time series data were extracted with ImageJ BAR plug-in and analyzed using custom Python data analysis pipelines.  $\% \Delta F/F_0$  was calculated on a trial-

by-trial basis as:  $((F-F_0)/F_0) \times 100$ , where  $F_0$  is the average of the three seconds preceding the start of the CS presentation. To classify neurons based on their responses to CS and US (inhibited, not-responding or excited), we compared the area under the curve (AUC) during 1.5 s baseline versus the AUC during the 1.5 s anticipatory period (CS and delay) or US presentation (Wilcoxon Rank Sum Test, using a significance threshold of  $p \leq 0.05$ ).  $\chi^2$  test was then used to compare the distribution of responses in CS+ and CS- trials, and across sessions. We performed decoding analyses to determine if neuronal activity could be used to predict the general mouse eyeblink during cue presentation on any given trial. To do so, we used a binary decoder for each neuron where its activity (computed as AUC) during cue presentation (i.e. from 0 to 1.5 s from cue starting) was used to predict whether the eye blinking (as CR) was lower than the 60th percentile or higher than the 40th percentile of that particular session. To complete the decoding analysis, we used the Python module Scikitlearn with GridSearchCV and a support vector classification (SVC) estimator [327]. This included a radial basis function kernel. Quantification of performance was done using 100 folds validation for each neuron, the average accuracy score across these parameters was used as the metric of accuracy. In order to determine whether the neuron's accuracy scores across all repetitions were significantly different from that expected by chance, we performed a single shuffle per neuron by randomizing the cue identity on every trial and tested using an unpaired t-test.

### C.2.8 Fiber photometry recordings

Fiber photometry recordings were performed using the ChiSquare X2-200 system ( $\chi^2$  Biomaging, Brookline, MA). Briefly, blue light from a 473 nm picosecond-pulsed laser (at 50 MHz; pulse width  $\approx 80$  ps FWHM) was delivered via a single mode fiber. Fluorescence emission was collected by a multimode fiber with a sampling frequency of 100 Hz. The single mode and multimode fibers were arranged side by side in a ferrule that is connected to a detachable multimode fiber implant. The emitted photons were bandpass-filtered (FF01-550/88, Semrock) to a single photon detector. Photons were recorded by the time-correlated single photon counting (TCSPC) module (SPC-130EM, Becker and Hickl, GmbH, Berlin, Germany) in the ChiSquare X2-200 system. Photometric recordings were analyzed using a custom made python code where the signal was smoothed (running average: 10) and  $\% \Delta F/F_0$  was calculated on a trial-by-trial basis as:  $((F - F_0)/F_0) \times 100$ , where  $F_0$  is the average of the three seconds preceding the start of the CS (or US presentation for unpredicted US trials). To determine whether we were able to collect valid signal, for each mouse, we analyzed the averaged response during the habituation session to CS and US presentation. We kept only the mice where we could detect positive or negative photometric responses when the mean number of photons/bin in at least 3 epochs (50 ms per epoch) was higher than the baseline average plus two times the Standard Deviation (SD) or lower than baseline minus two times the SD.

### C.2.9 In vivo electrophysiological recordings

Juxtacellular recordings of single LHb neurons (dataset in [Figure 3.8](#)) were performed as previously described [83, 103]. A subset of recordings ( $n_{\text{neurons}}=149$ ) was obtained in animals anesthetized with a mix of ketamine/xylazine according to previously published procedures [103]. Briefly, craniotomies were performed at stereotaxic coordinates (1.5 mm posterior, 0.5 mm lateral to bregma) above the LHb. Glass electrodes (impedance: 7-9 MOhm) were filled with a tracer-containing solution (1.5–2% Neurobiotin, SP-1120, Vector Laboratories). Juxtacellular voltage signals were acquired (ELC-03XS, NPI Electronic, Tamm, Germany) and digitized at 25 kHz (POWER1401-3 and Spike2 v. 8.02e, CED, Cambridge, UK). Spontaneous spiking was monitored for at least 5 min, followed

by foot shock stimulation (1 mA, 0.5 s, every 5 s, number of trials=20–197 trials) to the contralateral hind limb (A365R Stimulus Isolator, WPI). Juxtacellular labeling was performed using standard procedures. Recordings that were histologically confirmed to be outside of the LHb and recordings containing more than one unit or indications of cell damage were excluded from the analysis. Spikes from juxtacellular signals were isolated through threshold-based peak detection and visual inspection assisted by PCA, essentially as previously described [59]. After recording, animals were euthanized with an overdose of pentobarbital, transcardially perfused and their brains processed for histological analysis as previously described [28] (see also ‘histology and immunohistochemistry’ above). Part of the juxtacellular dataset in [Figure 3.8](#) was previously obtained ( $n_{\text{neurons}}=121$ ) from part of the data included in Congiu et al. 2019 and Lecca et al. 2017 [215]. Experimental procedures were carried out as previously described [83, 215]. Briefly, mice were anesthetized using isoflurane (Induction: 4%; maintenance: 1–1.5%) and placed in the stereotaxic apparatus (Kopf, Germany). Their body temperature was maintained at  $36 \pm 1^\circ\text{C}$  using a feedback-controlled heating pad (CMA 450 Temperature controller, Phymep). The scalp was retracted, and one burr hole was drilled above the LHb (AP: 1.3 to 1.6 mm, L: 0.35–0.5 mm, V: 2.3 to 3.2 mm) for the placement of a recording electrode. Single-unit activity was recorded extracellularly using glass micropipettes filled with 2% Chicago sky blue dissolved in 0.5 M sodium acetate (impedance 5–15 M $\Omega$ ). The signal was filtered (band-pass 500–5000 Hz), pre-amplified (DAM80, WPI, Germany), amplified (Neurolog System, Digitimer, UK), and displayed on a digital storage oscilloscope (OX 530, Metrix, USA). Experiments were sampled on- and offline by a computer connected to CED Power 1401 laboratory interface (Cambridge Electronic Design, Cambridge, UK) running the Spike2 software (Cambridge Electronic Design). Single units were isolated, and the spontaneous activity was recorded for a minimum of 3 min before assessing their response to foot shock delivery (duration: 0.5 s, intensity: 1 mA, ITI: 5 s). At the end of each experiment, mice were euthanized (overdose of isoflurane prior to killing) and the electrode placement was determined with an iontophoretic deposit of pontamine sky blue dye (1 mA, continuous current for 5 min). Brains were then rapidly removed and fixed in 4 % paraformaldehyde solution. The position of the electrodes was identified with a microscope in coronal sections (60  $\mu\text{m}$ ). Only recordings in the correct area were considered for analysis.

### C.2.10 In vitro electrophysiology

Mice were anesthetized (ketamine/xylazine; 150 mg/100 mg kg<sup>-1</sup>), sacrificed, and their brains were transferred in ice-cold carbogenated (95% O<sub>2</sub>/ 5% CO<sub>2</sub>) solution, containing (in mM) choline chloride 110; glucose 25; NaHCO<sub>3</sub> 25; MgCl<sub>2</sub> 2; ascorbic acid 11.6; sodium pyruvate 3.1; KCl 2.5; NaH<sub>2</sub>PO<sub>4</sub> 1.25; CaCl<sub>2</sub> 20.5. Coronal brain slices (250  $\mu\text{m}$  thickness) were prepared and transferred for 5 min to a warmed solution (34 °C) of identical composition, before transfer at room temperature in a carbogenated artificial cerebrospinal fluid (ACSF) containing (in mM) NaCl 124; NaHCO<sub>3</sub> 26.2; glucose 11; KCl 2.5; CaCl<sub>2</sub> 2.5; MgCl<sub>2</sub> 1.3; NaH<sub>2</sub>PO<sub>4</sub> 1. During recordings, slices were continuously superfused with ACSF at a flow rate of 2.5 mL/min at 32 °C. Neurons were patch-clamped using borosilicate glass pipettes (2.7–4 M $\Omega$ ; Phymep, France) under an Olympus-BX51 microscope (Olympus, France). The signal was amplified, filtered at 5 kHz and digitized at 10 kHz (Multiclamp 200B; Molecular Devices, USA). Data were acquired using Igor Pro with NIDAQ tools (Wavemetrics, USA). Access resistance was continuously monitored with a  $-4$  mV step delivered at 0.1 Hz. Extracellular stimulation from AMPI ISO-Flex stimulator was delivered through glass electrodes placed in the LHb. All recordings were made in voltage-clamp configuration. For AMPA/GABA evoked excitatory currents were recorded at  $-60$  mV, and evoked inhibitory currents were recorded at  $+5$  mV. AMPA/NMDA ratios were obtained recording compound AMPA+NMDA EP-

SCs at +40 mV, and subtraction of the APV-insensitive component (APV, 100  $\mu$ M). The paired-pulse ratio was obtained by recording the neurons at  $-60$  mV and by delivering two pulses at 50 msec interpulse intervals. Spontaneous excitatory currents were pharmacologically isolated by bath application of picrotoxin (PTX, GABAAR antagonist; 100  $\mu$ M). The internal solution contained (in mM) CsMeSO<sub>3</sub> 120, CsCl 10, HEPES 10, EGTA 10, creatine phosphate 5; Na<sub>2</sub>ATP 4; Na<sub>3</sub>GTP 0.4, QX-314 5. All drugs were purchased from HelloBio.

### C.2.11 Real time place aversion/preference test

C57BL/6 J wild-type male mice 8–10 weeks old were tested in a custom- made behavioral arena with two compartments with different visual cues and wall texture for 15 min. The time spent on each compartment was recorded via a digital camera interfaced with Ethovision software (Noldus). Paired-side for the CS+ auditory tone was randomly assigned and balanced among experimental subjects and groups. Mice were habituated for one session to the tone to avoid novelty-induced effects. At the beginning of the test session, the mouse was placed on the non-paired side of the chamber. Every time the mouse crossed to the paired side CS+ tone was played (duration 1 s; ITI 5 s) via a hardware timed-signals controlled by an I/O box (Noldus). The sound was interrupted at any instance at which the mouse left the paired side.

### C.2.12 Quantification and statistical analysis

All statistical analyses were conducted using Prism (v.9, GraphPad) or the Python scikit-learn library. Statistical tests used in this study include paired and unpaired t-test, Wilcoxon matched-pairs tests, Chi-Square test, one-way and two-way analysis of variance (ANOVA). When parametric tests were used, data normality was confirmed using the Shapiro–Walk normality test. P values were corrected for multiple comparisons when necessary. In the box plots, the center lines indicate the median, and the box limits indicate the upper and lower quantiles (95%); box plots include single observations. The significance threshold was held at  $\alpha = 0.05$  and tests were two-sided. Sample sizes were not predetermined using statistical methods but based on experimental settings within the laboratory. Experiments were randomized whenever possible. Experimenters were not blind to the experimental group apart from the in vitro electrophysiological recordings. Experiments were replicated at least two times within the laboratory. t-SNE embeddings were generated using 15 electrophysiological features: the first four principal components of interspike interval distributions (ISIs; bin size: 10 ms; computed for a range of 1 s), the first four principal components of autocorrelograms (ACGs; bin size: 1 ms; computed for a range of 500 ms and a narrower range of 100 ms), spontaneous firing rate, coefficient of variation (CV; ISI standard deviation/mean ISI, as in Softky and Koch 1993), and burst index (fraction of ISIs smaller than 25 ms). To classify neurons as footshock excited, inhibited, and non-modulated, we compared the baseline firing rate ( $FR_{baseline}$ ; 1 s prior to stimulus onset) with the stimulus firing rate ( $FR_{stimulus}$ ; 1 s after stimulus onset). Out of the 220 neurons with footshock stimulation, only neurons with a >20 footshock stimulating trials ( $n_{neurons} = 196$ ) were included in this analysis, in order to have adequate statistical power. Neurons were classified as FS-excited or FS-inhibited by comparing  $FR_{stimulus}$  and  $FR_{baseline}$  (Wilcoxon rank sum test,  $\alpha = 0.5$ ); neurons with  $FR_{stimulus} > FR_{baseline}$  were assigned to the FS-excited group and neurons with  $FR_{stimulus} < FR_{baseline}$  were instead assigned to the FS-inhibited group. To quantify the strength of foot shock modulation, we computed a foot shock modulation index (FS-modulation) defined as:  $(FR_{stimulus} - FR_{baseline}) / (FR_{stimulus} + FR_{baseline})$ .

## D | Supplementary material for Study IV: Zouridis et al. (2024)

### **D.1 Supplementary Figures**

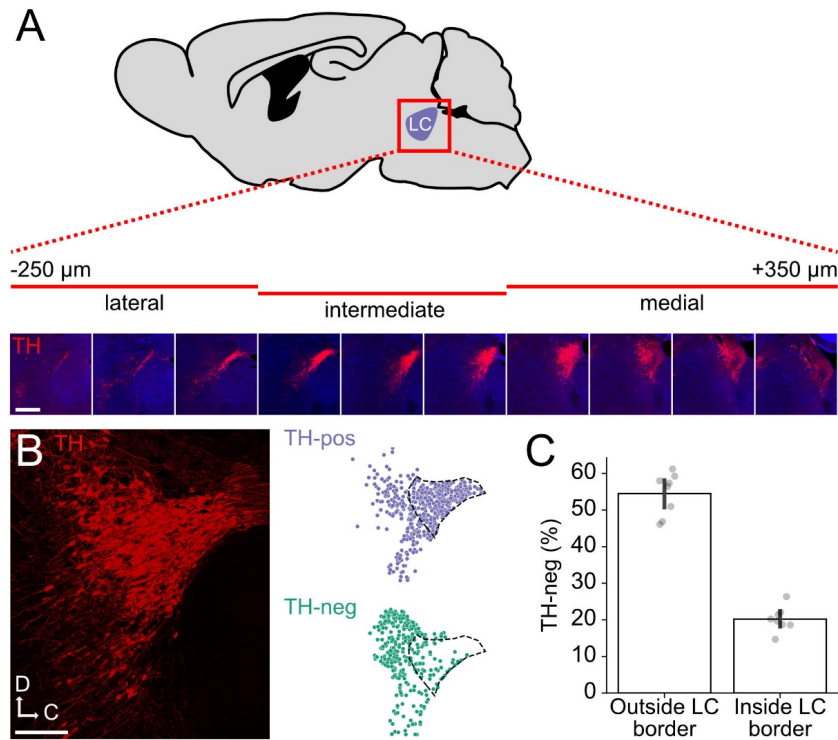


**Figure D.1: Histological analysis of the mouse LC complex.** (A) Top: Schematic of a parasagittal section of the mouse brain through the LC complex (magenta). Bottom: Sequential tissue sections immunostained for tyrosine hydroxylase (TH; red) at different mediolateral levels (lateral, intermediate, medial) of the LC complex. The intermediate level contains the dense grouping of noradrenergic cells, classically referred to as the LC 'nuclear core'. Tissue section thickness, 50  $\mu\text{m}$ ; scale bar, 200  $\mu\text{m}$ . Same orientation as in the schematic, i.e., top: dorsal, right: caudal. (B) Left: Parasagittal tissue section through the LC complex stained for TH (red). Scale bar, 200  $\mu\text{m}$ ; D, dorsal; C, caudal. Right: Corresponding annotation of TH-positive (magenta) and TH-negative (green) neurons using TH-NeuN-DAPI co-staining (see Methods). The posterior-dorsal aspect of the LC complex with a high packing density of TH-positive neurons is denoted by a dashed line. (C) Quantification of the relative proportion of TH-negative neurons inside and outside the posterior-dorsal aspect of the LC complex, as per the boundary shown in B ( $54.52 \pm 5.41\%$  TH-negative cells outside versus  $20.20 \pm 3.13\%$  inside LC border; 4933 neurons from  $n=8$  sections, 3 nuclei, 2 animals). Note that cell counting along the anteroposterior LC axis was performed to visually show that away from the LC 'nuclear core', the number of noradrenergic neurons decreases, and the percentage of non-TH cells increases accordingly (the same logic applies to all other axes away from the LC 'nuclear core').

D. Supplementary material for Study IV:  
Zouridis et al. (2024)

---

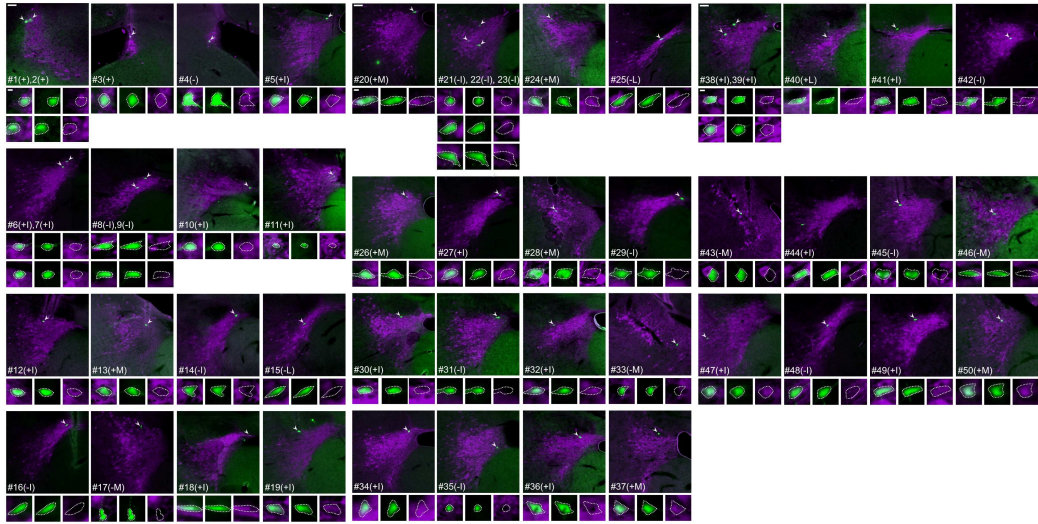


Figure D.2: **Histologically identified neurons within the LC complex.** Top panels: Low magnification fluorescence images stained for tyrosine hydroxylase (TH; magenta) and Neurobiotin (Nb; green) showing the anatomical location of the identified neurons. Each subpanel corresponds to one or more neurons labeled in a single LC complex. Section orientation: coronal (for cells #1-4) and parasagittal (for cells #5-50; up=dorsal, right=caudal). Scale bar, 100  $\mu$ m. Indicated: Cell ID; Neurochemical Identity, TH-positive (+) or TH-negative (-); Mediolateral Level, Medial (M) / Intermediate (I) / Lateral (L) (for reference, see [Figure D.1A](#)). Lower panels: high-magnification images showing the colocalization of Nb and TH immunofluorescence signals to determine the neurochemical identity of the neuron (see details in Methods). Scale bar, 10  $\mu$ m.

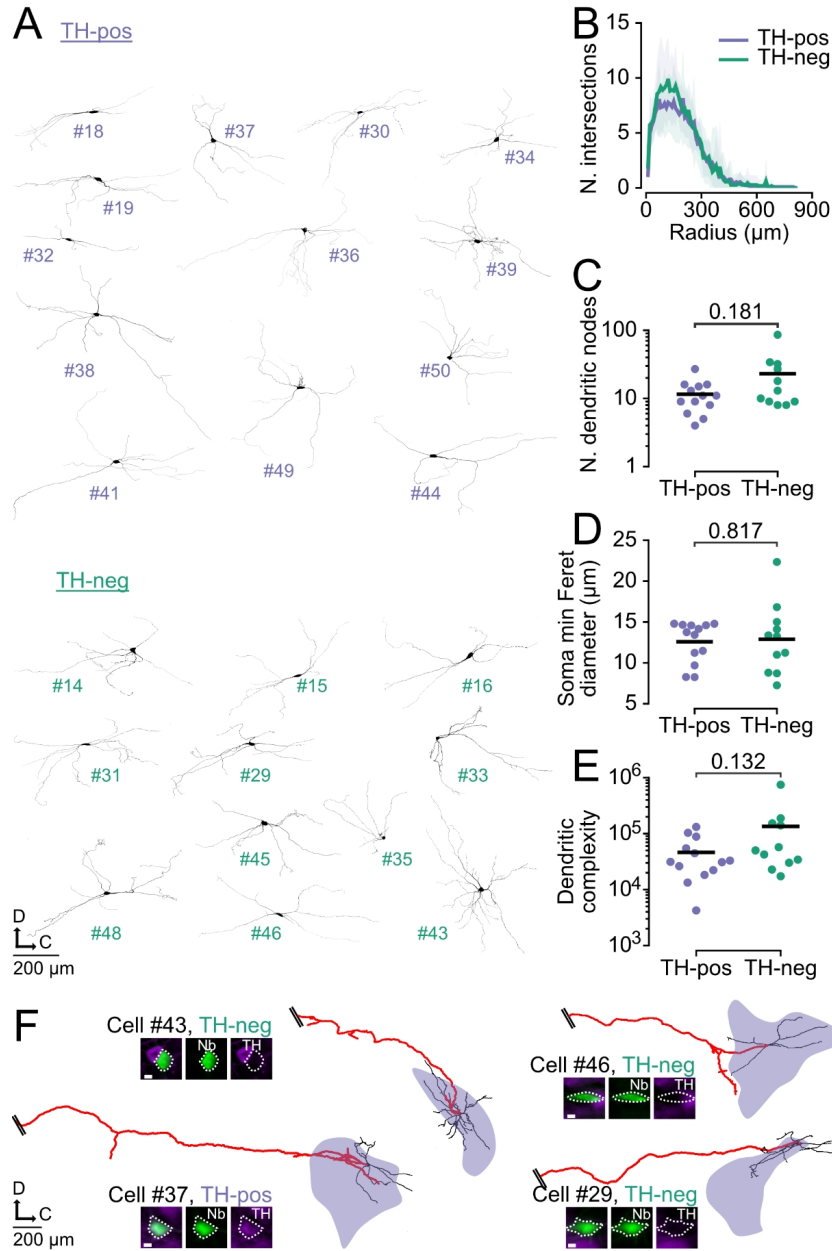


**Figure D.3: Morphological properties of TH-positive and TH-negative neurons.** (A) Morphological reconstructions of somatodendritic compartments of TH-positive (top, n=13) and TH-negative (bottom, n=11) neurons. Cell IDs are indicated (for corresponding epifluorescence microscopy images, see [Figure 3.14](#)). D, Dorsal; C, Caudal. (B-E) Morphometric somatodendritic features of TH-positive and TH-negative neurons. (B) Sholl intersection profiles, shaded regions indicate the standard deviation, (C) number of dendritic nodes, (D) soma diameter (estimated as the minimum Feret diameter), and (E) dendritic complexity (see details in Methods, n as in A). P-values are indicated (two-sided Mann-Whitney-Wilcoxon test, not corrected for multiple comparisons). (F) Morphological reconstructions of somatodendritic (black) and axonal (red) compartments of representative in vivo recorded and identified TH-positive (magenta) and TH-negative neurons (green). The thickness of dendrites and axons has been adjusted for display purposes. Axons were partially traced until they exited the LC complex (see Methods). Fluorescence images show corresponding immunostainings for TH (magenta) and Neurobiotin (green). Scale bars, 10  $\mu$ m. Note that partial axonal reconstructions show one main axonal branch exiting the LC complex (magenta-shaded area) in a rostral direction. D, Dorsal; C, Caudal.

D. Supplementary material for Study IV:  
Zouridis et al. (2024)

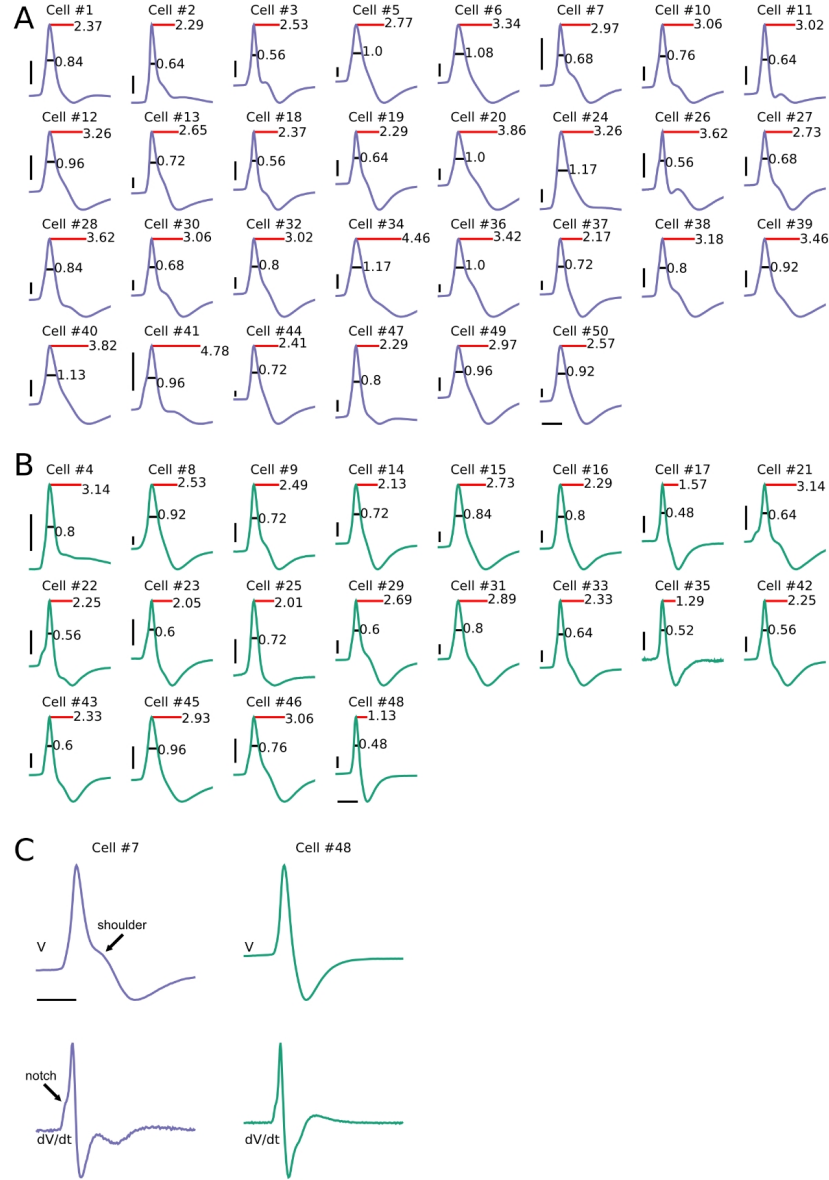


Figure D.4: **Spike waveforms of the identified neurons in the mouse LC complex.** (A) Mean spike waveforms for TH-positive ( $n=30$ ) neurons. Peak-to-trough (red line) and spike half-width (black line) durations are indicated (ms). Scale bars, 1 mV, 2 ms. (B) Same as in A, for TH-negative neurons ( $n=20$ ). (C) Top panels: normalized mean juxtacellular spike waveform for a representative TH-positive (Cell #7; magenta) and a representative TH-negative neuron (Cell #48; green). Bottom panels: corresponding first-order time-derivative of the spike waveform shown above. The first-order time derivative of the juxtacellular spike waveform has been shown to approximate the extracellularly recorded spike waveform. Note the presence of a 'notch' in the ascending phase and a 'shoulder' in the descending phase of the TH-positive spike waveform (indicated with arrows), which is consistent with previous reports (for comparison, see Fig. 3 in [395]; Fig. 1 in [438]). Scale bar, 2 ms.

D. Supplementary material for Study IV:  
Zouridis et al. (2024)

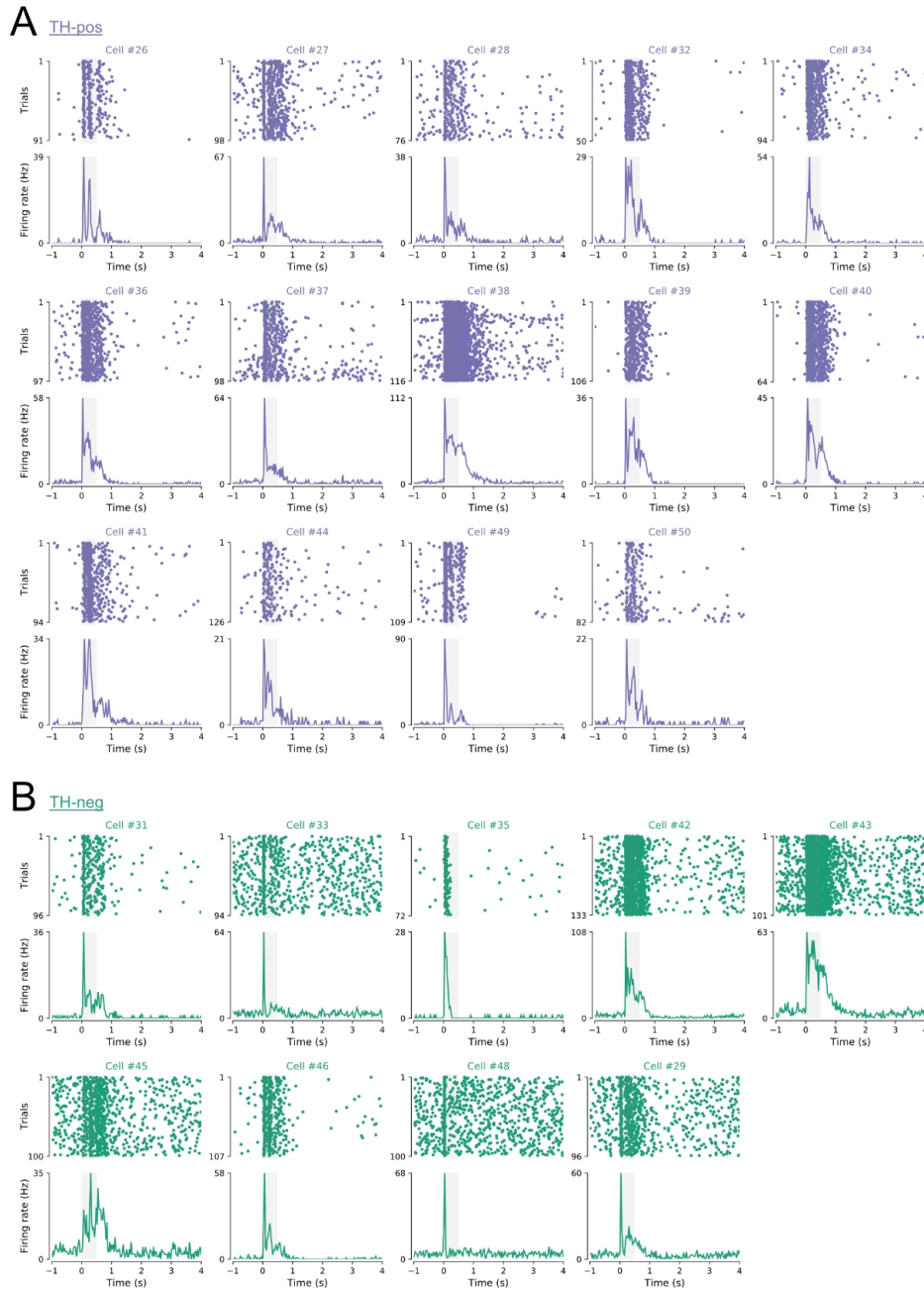


Figure D.5: **Footshock responses of identified neurons in the mouse LC complex.** Spike raster plot (top panels) and peristimulus time histogram (respective bottom panels) during footshock stimulation (gray-shaded area) for n=14 TH-positive (A; magenta) and n=9 TH-negative (B; green) neurons.

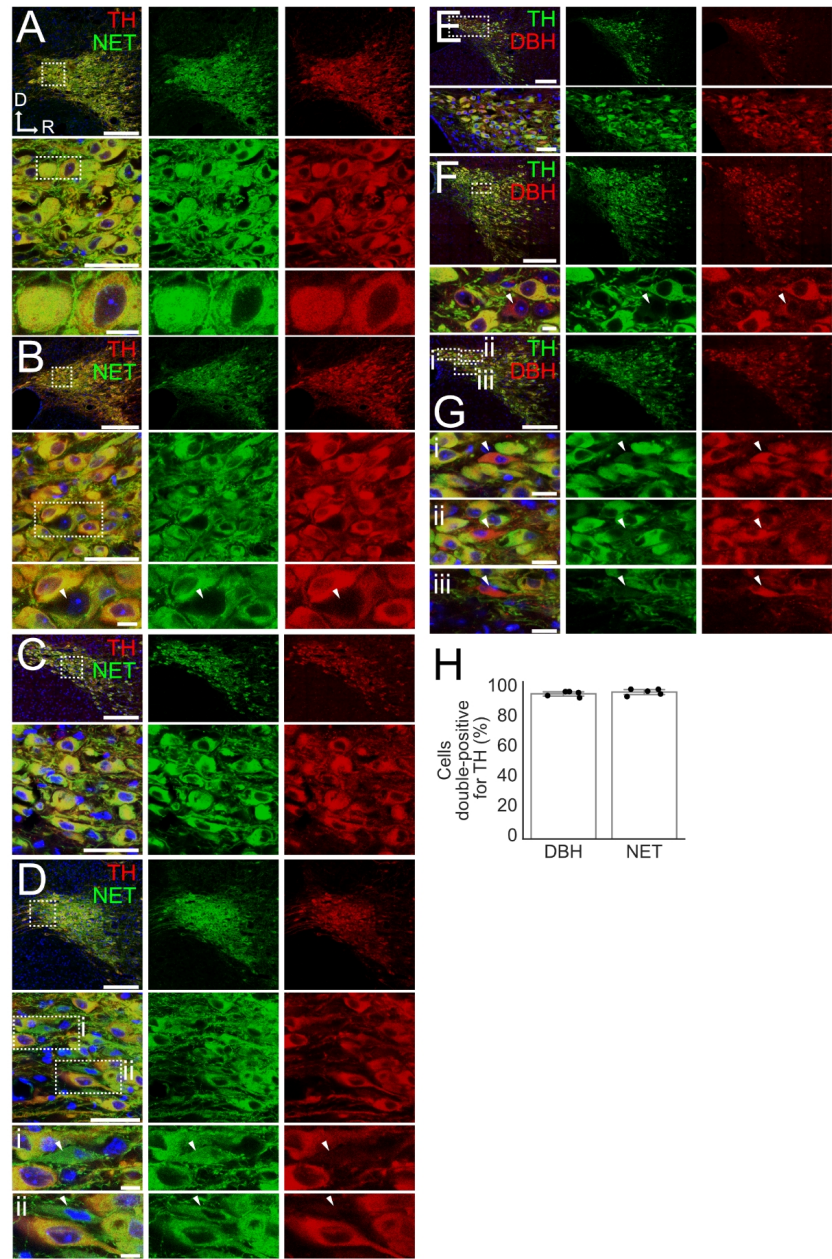


Figure D.6: **Colocalization of TH with additional noradrenergic markers (DBH and NET).** (A-D) Representative parasagittal tissue sections through 4 mouse LC complexes (from 3 animals), immunostained for tyrosine hydroxylase (TH; green), dopamine beta-hydroxylase (DBH; red), and DAPI (blue). Top: low-magnification view. Bottom: high-magnification details of the indicated areas (dotted boxes). Note the high degree of colocalization between the TH and DBH signals. A minority of neurons showed comparatively weaker TH signal intensity (indicated with arrowheads in panels D i and ii). For stringency, these weakly positive TH-neurons were considered TH-negative in the quantification (see panel H below). Considering these neurons as TH-positive resulted in 100% colocalization between TH and DBH (not shown). No DBH-negative or norepinephrine transporter (NET)-negative cells were found within the TH-positive population. (E-G) Same as in A-D but for TH (green) and NET (red). A minority of neurons with a comparatively weaker TH signal intensity (indicated with arrowheads in panels F and G) were considered TH-negative in the quantification (see panel H below). (H) Quantification of the degree of colocalization between TH with DBH or NET. DBH:  $96.95 \pm 1.83\%$  of TH-positive neurons also expressed the marker DBH (1614 neurons from  $n=5$  sections, 2 nuclei, 2 animals). NET:  $98.15 \pm 2.12\%$  of TH-positive neurons also expressed the marker NET (1489 neurons from  $n=5$  sections, 3 nuclei, 3 animals).

## D.2 Materials and methods

### D.2.1 Experimental animals

Experiments were conducted on adult C57BL/6J mice (male, >8 weeks old, ~20–30 g body weight; Charles River, Sulzfeld, Germany). Mice were housed in a 12 h light/dark cycle and provided ad libitum food and water. All experimental procedures were performed under the German Animal Welfare Act (TierSchG) and Animal Welfare Laboratory Animal Ordinance (TierSchVersV). The study was approved by the state authority (Regierungspräsidium, Tübingen, Baden-Württemberg, Germany; Licenses CIN7/19G, CIN5/19G, CIN1/19M, CIN1-2/22M).

### D.2.2 Surgical procedures

Surgical procedures were essentially performed as previously described [28, 106? ]. Animals were anaesthetized with a mixture of ketamine and xylazine (i.p., 80–100 mg/kg and 10 mg/kg, respectively), and the anaesthesia was maintained with additional doses of ketamine (half of the initial dose). The animal's body temperature was kept at 37 °C using a feedback-controlled heating pad (DC temperature controller, FHC). Ophthalmic ointment was applied to the eyes. Prior to the scalp incision, local anaesthesia was applied (s.c., Xylocaine, AstraZeneca, GmbH, Wedel, Germany). Craniotomies were performed above the LC (for the specific stereotaxic coordinates, see Section D.2.3).

### D.2.3 Juxtacellular recording and labelling

Micropipette electrodes were pulled from borosilicate capillaries (Cat#1403547, Hilgenberg, Malsfeld, Germany; Micropipette Puller Cat#P-1000, Sutter Instrument, Novato, CA, USA) to a ~1.5- $\mu$ m tip diameter (resistance 7–9 M $\Omega$  in Ringer's solution) and filled with a tracer-containing solution (1.5%–2% Neurobiotin, SP-1120, Vector Laboratories) in 'intracellular-like' solution (in mM): 135 K-gluconate, 10 HEPES, 10 Na2-phosphocreatine, 4 KCl, 4 MgATP and 0.3 Na3GTP, osmolarity: 280–310 mOsm). In our and others' observations, 'intracellular-like' and 'extracellular-like' electrode solutions do not significantly alter cells' spontaneous firing even during lengthy stimulation or entrainment procedures [102, 169, 170], likely due to their limited diffusion in the absence of applied air pressure. Broadband (high-pass, DC to 1 Hz; low-pass, 9 kHz) juxtacellular voltage signal was obtained via an ELC headstage, amplified 1000 times (ELC-03XS, NPI Electronic, Tamm, Germany) and digitized at either 25 or 50 kHz (POWER1401-3, Spike2 v. 8.02e, CED, Cambridge, UK). The electrode was advanced towards the LC (stereotaxic coordinates: 5.2–5.4 mm posterior, 0.85–0.95 mm lateral to bregma, 2.8–3.1 mm ventral to the brain surface) under electrophysiological control, and the impedance was monitored by applying negative current 'search' pulses (current intensity, 1 nA; duration, 200 ms; every 3 s). Upon establishing a juxtacellular recording configuration within the target area, tentative identification of LC neurons was based on their low spontaneous discharge rates (typically <5 Hz) and broad spike waveforms (e.g., [2, 395, 405, 438]). When advancing the electrode (vertical penetrations), juxtacellular recordings were often successfully established while entering into the LC complex, which possibly resulted in a sampling bias towards its dorsal border (Figure D.2). Spontaneous firing of neurons was recorded for  $289.8 \pm 83.37$  s prior to initiating either juxtacellular labelling directly or prior to initiating the foot shock stimulation protocol. In a subset of recordings ( $n = 23$ ), foot shock stimulation ( $95.81 \pm 19.32$  trials, range: 50–133) was performed by delivering square current pulses (current intensity, 1 mA, stimulus duration, 0.5 s; inter-stimulus

interval, 5 s) to the contralateral hind limb through a constant current stimulus isolator (Cat#A365R, WPI, Sarasota, FL, USA). Juxtacellular labelling was initiated with a brief, ‘buzz-like’ current pulse [28, 102] and carried out with positive square current pulses (200 ms, 50% duty cycle as in Pinault, 1996). During ‘entrainment’, the current intensity was manually adjusted, typically in the 1–20 nA range, to modulate spiking activity without damaging the cell. Following a post-labelling survival time of 1–3 h under deep general anaesthesia, histological analysis was performed (see D.2.4).

### D.2.4 Histological analysis, morphological reconstruction and cell quantification

Animals were euthanized with an overdose of pentobarbital (i.p., >300 mg/kg; Narcoren, Boehringer Ingelheim Vetmedica GmbH, Ingelheim, Germany) and transcardially perfused with 0.1 M phosphate-buffered saline (PBS), followed by fixative (4% paraformaldehyde solution in 0.1 M PBS). The brain tissue was collected, incubated in fixative overnight and stored in 0.1 M phosphate buffer (PB) at 4°C. Following cryoprotection (successive overnight incubations in 15%, 30% sucrose [Cat#4621.1, Carl Roth GmbH, Karlsruhe, Germany] in PBS) and freezing (-20°C, in blocks of Surgipath FSC 22 Clear, Leica Biosystems, Richmond, IL, USA), brain tissue was sliced on a cryostat (Cat#CM3050 S, Leica Microsystems, Germany) to obtain thin (70 µm for morphological reconstruction and neuronal tracing, 50 µm for cell quantification experiments) parasagittal sections (with the exception of four juxtacellularly labelled cells from two animals which were processed on coronal sections; see cells #1–4 in Figure D.2). To visualize the Neurobiotin-labelled cells, brain slices were incubated overnight at 4°C with a fluorescent streptavidin-Alexa Fluor 546 conjugate (1:1000, Cat#S11225, ThermoFisher Scientific, Waltham, Massachusetts) after permeabilization in PBS with 0.1% Triton X-100. The tissue sections were then incubated for 15 min with a 1:1000 DAPI solution (Cat#D1306, Thermo-Fisher Scientific, Waltham, Massachusetts) at room temperature, followed by three washes with PBS. Immunostainings were performed on free-floating tissue sections, essentially as previously described [28, 106]. The primary antibodies used were rabbit anti-TH (1:1000, Cat#P40101-150, Pel Freez Biologicals, as in [189]), mouse anti-TH (1:1000, Cat#MAB318, Sigma-Aldrich), mouse anti-NET (Norepinephrine Transporter; 1:1000, Cat#NET05-2, Mab Technologies), rabbit anti-DBH (Dopamine-beta-Hydroxylase; 1:500, Cat#22806, ImmunoStar) and rabbit anti-NeuN (1:500, Cat#MAB377, Merck Millipore). Secondary antibodies (1:400, Goat anti-Rabbit Secondary Antibody Alexa Fluor 488, Cat#A-11034, Thermo Fisher Scientific, and Goat anti-Mouse Secondary Antibody Alexa Fluor 488, Cat#A-21424, Thermo Fisher Scientific) were incubated at room temperature for 4 h. The sensitivity and specificity of the two anti-TH antibodies were virtually identical, as indicated by the 100% colocalization between the two stainings (639/639 double-positive neurons; not shown). Fluorescence images were acquired via epifluorescence (5×, 10×, 20×; Axio Imager Z1, Carl Zeiss, Oberkochen, Germany) or confocal microscopy (40×, 63×; LSM710, Carl Zeiss, Oberkochen, Germany). For each cell, the neurochemical assignment was assessed based on visual inspection of usually multiple serial z-stacks through the soma and confirmed by two independent experimenters. For colour-blind accessibility and display purposes, fluorescence images were converted into green and magenta, and intensities were linearly adjusted. To analyse cell morphologies, a horseradish peroxidase reaction was performed on Neurobiotin streptavidin-labelled sections, followed by conversion with a 3,30’-diaminobenzidine (DAB) substrate enhanced with Ni<sup>2+</sup> as previously described [198, 320]. Somatodendritic domains of neurons were manually reconstructed using Neurolucida (MBF Bioscience, Williston, Vermont). Long-range axons were only partially reconstructed (Figure D.3F; n = 4 neurons) and typically consisted of one main branch that exited the LC complex towards the rostral direction. A post hoc adjustment accounted for tissue shrinkage in the

Z dimension. The resulting dataset consisted of  $n = 24$  (TH-positive  $n = 13$ , TH-negative  $n = 11$ ) morphological reconstructions of neurons obtained from 17 animals. Quantitative morphological features of neuronal reconstructions were derived using the Neurolucida software (v. 2022.2.3). For dendritic structures (**Figure D.3A–E**), the number of nodes (branching points), Sholl intersection profiles (dendritic intersections quantified in concentric spheres of  $10\ \mu\text{m}$  radius centred at the soma) and dendritic complexity (see [307], defined as  $(\text{Sum of the terminal orders} + \text{number of terminals}) \cdot (\frac{\text{total dendritic length}}{\text{number of primary dendrites}})$ ), were calculated. Somatic shape and size (**Figure D.3D**) were assessed using the minimum and maximum Feret diameters (respectively, the smallest and largest dimensions of the soma contour, as measured in the reconstruction slice where the contour area was maximal). Neuronal quantification and colocalization analysis (**Figure 3.13b**, **Figure D.1B,C**, and **Figure D.6**; TH-NeuN:  $n = 7$  sections from 3 LC nuclei, 2 animals; TH-NET:  $n = 5$  sections from 3 LC nuclei, 3 animals; TH-DBH:  $n = 5$  sections from 2 LC nuclei, 2 animals) was performed on confocal microscopy images ( $40\times$ ,  $63\times$  magnification, z-stacks) using ImageJ [344] and the Colocalization Object Counter Plugin (Lunde & Glover, 2020). TH-positive cells were quantified using both TH and NeuN signals, as the latter tends to be less intense in TH-positive neurons (**Figure 3.13B**; [65]).

## D.2.5 Electrophysiology data analysis

Spikes were detected offline, essentially as previously described [? ]. Briefly, a thresholded local maxima detection method was applied to high-pass filtered (cut-off frequency, 300 Hz) voltage traces to identify candidate spikes. Visual inspection of spike waveforms and the first two principal components allowed the spikes to be reliably isolated from artefacts. Parts of recordings displaying evidence of cellular damage (specifically, spike broadening, increased discharge frequency, prolonged negative DC shift; see [155, 312]) were not included in further analysis. The resulting electrophysiology dataset consists of 50 recordings (in 32 animals) from anatomically identified LC neurons, 23 of which (in 14 animals) include foot shock stimulation. Multiple recordings were conducted (bilaterally: 21 recordings from 10 animals: cells #3–#4, #12–#13, #17–#19, #34–#35, #36–#37, #44–#45, #24–#47, #49–#50, #33–#48, #38; #39–#45; and unilaterally: 12 recordings from 5 animals: cells #1–#2, #5–#6–#7, #8–#9, #22–#23–#24, #38–#39; see **Figure D.2**). In experiments where multiple labelling attempts were performed within the same LC complex, recordings were only included in the dataset if the respective neurons could be identified. Three experiments yielded two neurons (cells #1, #2; cells #8, #9; cells #38, #39; see **Figure D.2**), and two experiments yielded three neurons (cells #5, #6, #7; cells #21, #22, #23; see **Figure D.2**). In two cases, it was unclear which recordings corresponded to which neurons (cells #1, 2; cells #5, 6, 7, see **Figure D.2**); however, since all neurons from each experiment were of the same class (i.e., TH-positive), the respective recordings were included in the analysis. Therefore, our electrophysiological dataset consists of 45 recordings from neurons identified both cytochemically and histologically and five recordings from neurons identified only cytochemically. The broadband juxtacellular voltage signal was used to extract spike waveforms (data snippets  $-2$  to  $+6$  ms aligned to the local maximum). After subtracting the DC offset for each spike waveform, the mean spike waveform was calculated. ‘Peak-to-trough’ (temporal difference between the prominent electropositive local maximum and the subsequent electronegative local minimum of a spike waveform) and ‘half-width’ (temporal difference between two values of a spike waveform at the half-maximal amplitude of the peak) durations were derived from the mean spike waveforms. The interspike interval (ISI) coefficient of variation (CV; calculated as the ratio of the standard deviation of the ISI over the mean ISI, as in Softky & Koch, 1993). Mean firing rates, mean spike waveforms, ISI, CV and autocorrelograms were computed based on baseline (spontaneous) firing, recorded prior to the initiation of juxtacellular labelling or foot shock stimulation. Temporal fea-

tures of the spike wave-forms, ISI distributions and autocorrelograms were extracted with sparse principal component analysis (sPCA; [463]) using the first four principal components (Figure 3.15A–C). Autocorrelograms were computed for the range  $\pm 0.5$  s and ISI distributions (Figure 3.15B–C) for 0–1 s, normalized by z-scoring. TH-positive and TH-negative neurons were compared using their projections on the first two principal components (Figure 3.15D–F). To assess neuronal activity related to foot shock stimuli, we constructed peri-stimulus time histograms ( $-0.5$  to  $+4.5$  s with respect to stimulus onset; bin width, 30 ms). To assess the neuronal response reliability to foot shock stimulation (Figure 3.16), we calculated the fraction of trials in which an evoked spike occurred within 100 ms of the foot shock onset. To determine the response latency to foot shock, we calculated the median latency of the first spikes occurring within 100 ms after stimulus onset. To quantify the strength of foot shock modulation, we computed a Response Modulation Index [? ], defined as  $\frac{FR_{stimulus} - FR_{baseline}}{FR_{stimulus} + FR_{baseline}}$ , and Response Ratio, defined as  $\frac{\max(FR_{stimulus})}{FR_{baseline}}$  where  $FR_{baseline}$  represents the mean firing rate during 1 s preceding the stimulus onset,  $FR_{stimulus}$  represents the mean firing rate during stimulus delivery and  $\max(FR_{stimulus})$  represents the peak response within 0–0.5 s of the peristimulus time histogram. To analyse the post-activation firing suppression by foot shock stimulation, a post-activation suppression index was calculated as  $\frac{FR_{iti} - FR_{baseline}}{FR_{iti} + FR_{baseline}}$ , where  $FR_{iti}$  represents the mean firing rate during  $+1.5$  to  $+4.5$  s post-stimulus onset, and  $FR_{baseline}$  represents the mean firing rate during spontaneous activity prior to the initiation of foot shock stimulation. To further validate our conclusions, we compared the electrophysiological properties of TH-positive and TH-negative cells in the following conditions:

1. Across different mediolateral levels of the LC (medial,  $n = 7$  TH-positive,  $n = 4$  TH-negative neurons; intermediate  $n = 19$  TH-positive,  $n = 13$  TH-negative neurons; lateral,  $n = 1$  TH-positive,  $n = 2$  TH-negative neurons;  $n = 3$  TH-positive and  $n = 1$  TH-negative neurons were not assigned to any mediolateral level due to histological assessment performed on coronal sections).
2. Within versus outside the posterior-dorsal aspect of the LC complex (Figure D.1B,C), which contains the highest density of noradrenergic cells ('inside,'  $n = 17$  TH-positive,  $n = 13$  TH-negative neurons; 'outside,'  $n = 13$  TH-positive and  $n = 7$  TH-negative neurons).
3. By restricting the analysis to neurons with 'high confidence' of immunohistochemical classification (based on the overall quality and intensity of the fluorescence signal; 'high confidence,'  $n = 27$  TH-positive and  $n = 12$  TH-negative neurons; 'low confidence,'  $n = 3$  TH-positive,  $n = 8$  TH-negative neurons).

In all the above cases, the differences in firing rate, spike half-width and peak-to-trough duration between TH-positive and TH-negative cells trended in the same direction as in our cumulative dataset (as in Figure 3.14a,c; data not shown), thus indicating that our conclusions are unlikely to be significantly biased by unintentional anatomical biases and/or cell type misclassification. We note that potential inadvertent misclassification would likely diminish the observed differences between TH-positive and TH-negative neurons.

## D.2.6 PCA embedding

To embed neurons in two-dimensional space, we performed PCA on the standardized matrix that has the number of neurons ( $n = 50$ ) as rows and the electrophysiological features as columns (Figure 3.15D). The arrows in the PCA biplot (Figure 3.15F) are obtained by  $V S / \sqrt{n - 1}$  with  $V$  as the eigenvector,  $S$  as the eigenvalues and  $n$  as the number of data

points. Showing the PCA loadings has the advantage of making the relations between features interpretable: the cosine of products between any two arrows approximates the covariance between them.

### D.2.7 Logistic regression

We used LASSO logistic regression (Tibshirani, 1996) to classify neurons into TH-positive and TH-negative classes. We included 17 features ([Figure 3.15D](#), [Figure 3.15F](#), [Figure 3.15G](#)) and used a cross-validated grid search to identify the optimal regularization ( $c = 1.21$ , inverse regularization strength). The data set was split into 80% training and 20% testing, and we used five folds for the cross-validation. ‘Accuracy’ was defined as the fraction of correct predictions (including both true positives and true negatives) among the total number of cases examined, as implemented in the Python library scikit-learn [303]. We note that the classifier approach was primarily intended as a confirmatory approach to the physiological differences observed between the two neuronal classes. The classifier has overall limited accuracy (76%) and was trained on a limited number of observations ( $n = 30$  TH-positive,  $n = 20$  TH-negative); hence, its applicability to other datasets and/or experimental preparations is likely to be limited. Foot shock response metrics were not included in this analysis, given the even more limited number of observations ( $n = 14$  TH-positive,  $n = 9$  TH-negative).

### D.2.8 Statistical analysis

Statistical analysis of pairwise comparisons was performed with the two-sided Mann-Whitney-Wilcoxon test. P-values were not adjusted for multiple comparisons. This is an exploratory study, and p-values should be interpreted accordingly. Results are reported as mean  $\pm$  sample standard deviation unless otherwise specified.