THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
dans le cadre d'une cotutelle avec l'Université de Montréal
le 8 février 2024 par

Hugo J. LADRET

A multiscale model to account for orientation selectivity in natural images

Discipline
Biologie Santé

Spécialité
Neurosciences

École doctorale
ED 62 - Sciences de la Vie et de la Santé

Laboratoire/Partenaires de recherche
Institut de Neurosciences de la Timone,
CNRS UMR 7289 et Aix-Marseille Université

Laboratoire des Neurosciences de la Vision,
Université de Montréal

Composition du jury

Elvire VAUCHER Université de Montréal	Présidente du jury
Laura DUGUÉ Université Paris Cité	Rapporteuse
Nicholas PRIEBE University of Texas at Austin	Rapporteur
Olivier MARRE Institut de la Vision	Examinateur
Jason ESHRAGHIAN University of California Santa Cruz	Examinateur·e
Lyle MULLER University of Western Ontario	Examinateur
Ede RANCZ Aix-Marseille Université	Membre invité
Laurent PERRINET Aix-Marseille Université	Directeur de thèse
Christian CASANOVA Université de Montréal	Co-Directeur de thèse

Affidavit

I, undersigned, Hugo J. Ladret, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific supervision of Laurent Perrinet and Christian Casanova in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the french national charter for Research Integrity and the Aix-Marseille University charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Marseille, 12th November 2023

À mes parents,
à ma famille,
à mes amis,
au grand Hasard de la vie.

Liste de publications et/ou brevets et participation aux conférences

Liste des publications et/ou brevet réalisés dans le cadre du projet de thèse :

1. **Ladret, H. J.**, Casanova, C., & Perrinet, L. U., (2023). *Kernel Heterogeneity Improves Sparseness of Natural Images Representations*. preprint available as an arXiv preprint arXiv :2312.14685, under review (as of 01/01/2024) in Neuromorphic Computing and Engineering.

2. Cortes, N., **Ladret, H. J.**, Abbas Farishta, R., & Casanova, C. (2023). *The pulvinar as a hub of visual processing and cortical integration*. published in Trends in Neurosciences, dec 2023.
   DOI :https ://doi.org/10.1016/j.tins.2023.11.008

3. **Ladret, H. J.**, Cortes, N., Ikan, L., Chavane, F., Casanova, C., & Perrinet, L. U. (2023). *Cortical recurrence supports resilience to sensory variance in the primary visual cortex*. published in Communications Biology, 6(1), 667.
   DOI :https ://doi.org/10.1038/s42003-023-05042-3

4. Cortes, N., Abbas Farishta, R., **Ladret, H. J.**, & Casanova, C. (2021). *Corticothalamic Projections Gate Alpha Rhythms in the Pulvinar*. published in Frontiers in Cellular Neuroscience, 15, 787170.
   DOI :https ://doi.org/10.3389/fncel.2021.787170

Participation aux conférences et écoles d'été au cours de la période de thèse :

Articles de conférences :

1. **Ladret, H. J.**, Perrinet, L. U., & Casanova, C. (2023). *Convolutional Sparse Coding is improved by heterogeneous uncertainty modeling*. Proceedings of ICLR SNN, Kigali, Rwanda.

2. **Ladret, H. J.**, Cortes, N., Ikan, I., Chavane, F., Casanova, C., & Perrinet, L. U. (2023). *Resilience to sensory uncertainty in the primary visual cortex*. Proceedings of Cosyne, Montreal, Canada.

Présentations orales :

1. **Ladret, H. J.** (2022). *Making sense of the visual mess.* Presented at Aix-Marseille's doctoral schools summit, Marseille, France.

2. **Ladret, H. J.** (2022). *Dealing with sensory variance in the primary visual cortex.* Invited talk at Netherlands Institute of Neuroscience, Amsterdam, Netherlands.

3. **Ladret, H. J.**, Perrinet, L. U. (2022). *Statistics of the sparse representations of natural images.* Presented at SIAM Conference, Virtual.

4. **Ladret, H. J.** (2021). *Modulation of orientation selectivity by orientation precision in V1.* Presented at GDR Vision, Lille, France.

5. **Ladret, H. J.**, Perrinet, L. U. (2021). *Dynamics of the processing of orientation precision in the primary visual cortex.* Invited talk at DynamicsDays, Sophia-Antipolis.

Présentations sous forme de poster :

1. **Ladret, H. J.**, Cortes, N., Ikan, I., Chavane, F., Casanova, C. & Perrinet, L. U. (2023). *Computing sensory variance through intracortical recurrence.* Poster session presented at VSS, Tampa, USA.

2. **Ladret, H. J.**, Cortes, N., Ikan, I., Chavane, F., Casanova, C. & Perrinet, L. U. (2023). *Resilience to sensory uncertainty in the primary visual cortex.* Poster session presented at Cosyne, Montreal, Canada.

3. **Ladret, H. J.** & Perrinet, L. U. (2022). *Uncertainty in, uncertainty out : epistemic variance improves encoding of natural images.* Poster session presented at GDR Vision, Toulouse, France.

4. Ikan, I., Cortes, N., **Ladret, H. J.**, Laplante, I. & Casanova, C. (2022). *Dynamics of response's accuracy in the visual cortical area 21a.* Poster session presented at SfN, San Diego, USA.

5. **Ladret, H. J.**, Cortes, N., Ikan, I., Chavane, F., Casanova, C. & Perrinet, L. U. (2022). *Input variance and V1.* Poster session presented at INT X Anniversary, Marseille, France.

6. Bourquin, N., Perrot, A., Porée, L., Belgharbi, S., Cortes, N., Miquel, C., Bélanger, E., **Ladret, H. J.**, Ikan, I., Thorin, C., Lesage, F. & Provost, J. (2022). *Dynamic Ultrasound Localisation Microscopy Achieves Quantitative pulsatility Measurements in the Whole Brain Using Kalman Filtering.* Poster session presented at IEEE US, Virtual.

7. **Ladret, H. J.**, Cortes, N., Ikan, I., Chavane, F., Casanova, C. & Perrinet, L. U. (2022). *Recurrent cortical connectivity in the primary visual cortex supports robust encoding of natural sensory inputs.* Poster session presented at FENS, Paris, France.

8. **Ladret, H. J.**, & Perrinet, L. U. (2022). *A resilient neural code in V1 to process natural images.* Poster session presented at AREADNE, Santorini, Greece.

9. **Ladret, H. J.**, Cortes, N., Ikan, I., Chavane, F., Casanova, C. & Perrinet, L. U. (2021). *Modulation of orientation selectivity by orientation precision.* Poster session presented at SfN, Virtual.

10. **Ladret, H. J.**, & Perrinet, L. U. (2021). *Decoding orientation distributions from noisy observations in the primary visual cortex.* Poster session presented at Champalimaud's Dialogues on Neural and Machine Intelligence, Lisbon, Portugal.

11. **Ladret, H. J.**, Cortes, N., Ikan, I., Chavane, F., Casanova, C. & Perrinet, L. U. (2021). *Processing of orientation precision in the primary visual cortex.* Poster session presented at NeuroFrance Annual Meeting, Strasbourg, France.

12. **Ladret, H. J.**, Cortes, N., Perrinet, L. U. & Casanova, C. (2019). *Orientation selectivity to synthetic natural patterns in a cortical-like model of the cat primary visual cortex.* Poster session presented at SfN, Chicago, USA.

13. **Ladret, H. J.** & Ibos, G. (2019). *Comparative decision-making in the Posterior Parietal Cortex : proof of concept.* Poster session presented at SfN, Chicago, USA.

14. **Ladret, H. J.**, Cortes, N., Casanova, C., & Perrinet, L. U. (2019). *Learning dynamics in a neural network model of the primary visual cortex.* Poster session presented at Vision Health Research Network, Quebec, Canada.

15. **Ladret, H. J.** & Perrinet, L. U. (2018). *Selectivity to oriented patterns of different precisions.* Poster session presented at GDR Vision, Paris, France.

Écoles d'été :

1. (2023) *Telluride Neuromorphic Cognition Engineering Workshop.* Project leader on a brain organoid / open-source software interface project. Telluride, Colorado, USA.

2. (2022) *European Summer School of Vision.* Schloss Rauischholzhausen, Hessia, Germany.

Résumé

Cette thèse vise à comprendre les fondements et les fonctions des calculs probabilistes impliqués dans les processus visuels. Nous nous appuyons sur une double stratégie, qui implique le développement de modèles dans le cadre du codage prédictif selon le principe de l'énergie libre. Ces modèles servent à définir des hypothèses claires sur la fonction neuronale, qui sont testées à l'aide d'enregistrements extracellulaires du cortex visuel primaire. Cette région du cerveau est principalement impliquée dans les calculs sur les unités élémentaires des entrées visuelles naturelles, sous la forme de distributions d'orientations.

Ces distributions probabilistes, par nature, reposent sur le traitement de la moyenne et de la variance d'une entrée visuelle. Alors que les premières ont fait l'objet d'un examen neurobiologique approfondi, les secondes ont été largement négligées. Cette thèse vise à combler cette lacune.

Nous avançons l'idée que la connectivité récurrente intracorticale est parfaitement adaptée au traitement d'une telle variance d'entrées, et nos contributions à cette idée sont multiples. (1) Nous fournissons tout d'abord un examen informatique de la structure d'orientation des images naturelles et des stratégies d'encodage neuronal associées. Un modèle empirique clairsemé montre que le code neuronal optimal pour représenter les images naturelles s'appuie sur la variance de l'orientation pour améliorer l'efficacité, la performance et la résilience. (2) Cela ouvre la voie à une étude expérimentale des réponses neurales dans le cortex visuel primaire du chat à des stimuli multivariés. Nous découvrons de nouveaux types de neurones fonctionnels, dépendants de la couche corticale, qui peuvent être liés à la connectivité récurrente. (3) Nous démontrons que ce traitement de la variance peut être compris comme un graphe dynamique pondéré conditionné par la variance sensorielle, en utilisant des enregistrements du cortex visuel primaire du macaque. (4) Enfin, nous soutenons l'existence de calculs de variance (prédictifs) en dehors du cortex visuel primaire, par l'intermédiaire du noyau pulvinar du thalamus. Cela ouvre la voie à des études sur les calculs de variance en tant que calculs neuronaux génériques soutenus par la récurrence dans l'ensemble du cortex.

Mots-clés : vision, variance, cerveau Bayesien, encodage prédictif, neurobiologie, neurocomputation, connectivité synaptique récurrente.

Abstract

"Dear Sir or Madam, will you read my book ?
It took me years to write, will you take a look ?"
The Beatles, Paperback Writer, 1966

This thesis aims to understand the foundations and functions of the probabilistic computations involved in visual processes. We leverage a two-fold strategy, which involves the development of models within the framework of predictive coding under the free energy principle. These models serve to define clear hypotheses of neuronal function, which are tested using extracellular recordings of the primary visual cortex. This brain region is predominantly involved in computations on the elementary units of natural visual inputs, in the form of distributions of oriented edges.

These probabilistic distributions, by nature, rely on processing both the mean and variance of a visual input. While the former have undergone extensive neurobiological scrutiny, the latter have been largely overlooked. This thesis aims to bridge this knowledge gap.

We put forward the notion that intracortical recurrent connectivity is optimally suited for processing such variance of inputs, and our contributions to this idea are multi-faceted. (1) We first provide a computational examination of the orientation structure of natural images and associated neural encoding strategies. An empirical sparse model shows that the optimal neural code for representing natural images relies on orientation variance for increased efficiency, performance, and resilience. (2) This paves the way for an experimental investigation of neural responses in the cat's primary visual cortex to multivariate stimuli. We uncover novel, cortical-layer-dependent, functional neuronal types that can be linked to recurrent connectivity. (3) We demonstrate that this variance processing can be understood as a dynamical weighted graph conditioned on sensory variance, using macaque primary visual cortex recordings. (4) Finally, we argue for the existence of (predictive) variance computations outside the primary visual cortex, through the Pulvinar nucleus of the thalamus. This paves the way for studies on variance computations as generic weighting of neural computations, supported by recurrence throughout the entire cortex.

Keywords: vision, variance, Bayesian brain, Predictive Coding, neurobiology, neuro-computation, recurrent cortical connectivity.

Acknowledgements

"En bande organisée, personne peut nous canaliser"
13'Organisé, Bande Organisée, 2020

The following pages recapitulate half a decade of exciting scientific adventures, a voyage brought to life by the indispensable contributions of many wonderful human beings.

First and foremost, this journey would not have been possible without the support of my two co-supervisors. Laurent, who many years ago embraced the idea of recruiting a molecular biologist to work on a computational project. Embraced it so well, actually, that he did not fire me when I googled "how to import NumPy" on my first day of work. Laurent, your constant presence throughout this journey has been an unwavering beacon, always showing the way for unconventional and bold ideas to pursue together. More than a supervisor, you are a mentor who has instilled in me a passion for the relentless pursuit of scientific discovery, and for that I am forever in your debt. Your generosity with the conference budget allowed to gather the massive "natural images dataset" used in the third chapter of this thesis; but also allowed us to foster many collaborations throughout this thesis, defining this adventure and many new ones to come.

Christian was one such collaborator, who then went on to become co-director of this journey. Just as Laurent took a leap of faith by allowing a biologist to embark on a computational project, you have doubled down on that courage by entrusting a newly converted computationalist with the reins of a biological project. Considering the tangible resources (and budget) at play, I'd say this shows a lot of your generous nature, as you welcomed me with open arms and unfaltering support, undeterred by (my many) experimental setbacks, (constant) requests for new computers, (unreasonable) new experimental materials or even (suspiciously frequent) new electrodes. The lessons of scientific independence and resilience learned during our adventure together will last a lifetime. Thank you also for letting me take home a lab's server during the COVID-19 lockdown, as I will now confess that the computational power unclaimed by my code might was also channeled into resource-intensive GPU applications (i.e. *Doom Eternal*) during these two interminable months. Rounding off my trio of navigators on this scholarly odyssey is Frédéric, who more than deserves the title of unofficial third supervisor. Fredo, I am grateful for your understanding and patience with my naïve questions regarding freshly harvested results, but also for your razor-sharp mind during their analysis, and keen eye when reviewing our papers (yes, I can already hear you say *"fayot !"* as you read these lines). I tried my best to not beat your record of "more paper cited than Fregnac", but I think I failed, although there might be one or

9

two that are in the citation list as jokes.

   The larger crew of this ship is composed of numerous individuals whose invaluable contributions have guided this expedition into the brain. Nelson, thank you for your tutelage in electrophysiology, which was nothing short of phenomenal, and for our *"sensei-professoro"* relationship that has evolved into one of the most cherished friendships of my life. Geneviève, you are the source of the magic that makes the Montreal lab alive, and I thank you for investing extra time to unveil the arcane workings of all our experimental tools. Lamyae, my companion in arms as we wrestled a capricious electrophysiology system, best wishes for your own thesis journey, and thank you for having single-handedly recorded 56% (I counted) of the neurons we used in the Nature Communications Biology article. I'm still amazed by your electrode-handling magic ! To the eminent scientists who played advisory roles in this thesis – David, Karim, Paul, and Steve – thank you for sharing the burden of administrative constraints with grace and fortitude, turning these hurdles into stepping stones on my path. Many thanks to the members of the jury, who have agreed to read a manuscript that is longer than it has any right to be. I could not think of anyone better suited than those who have gracefully agreed to sit on this defense, and count myself extremely lucky to have such eminent future "peers".

There is a rather lenghty list of people who contributed significantly to this thesis. Guillaume, thank you for a pinpoint accurate critique of a presentation I delivered years ago, which served as a wake-up call to catalyze a journey towards higher scientific standards. Guilhem, thank you for our (brief) collaboration in modeling the parietal cortex during the early stages of my PhD. Jonathan, thank you for having paved the way for all things related to the involvement of Motion Clouds in the world of neurophysiology. Antoine, your kindness, gourmet barbecues, and expert knowledge of logistic regression have made you a cherished colleague and friend. Jean-Nicolas, thank you for your refreshing sense of humor, our shared appreciation for what others might call disgusting stout beers, and for invigorating debates on deep learning. Patrice and Marion, for the many "apéros" and incredible luck of having you as friends, but also for endless exchanges of TP after one "apéro" too many. Louison, for walking through the callanques at hellish speeds, I owe you Figure 1 of Chapter 4's article ! On the mention of hiking, special thanks to Shashank, for having had the kindness of driving all the way to the end of Telluride's valley to pick me up, saving my toes and earning the right to name my firstborn (I hope you won't go too crazy). Justine, for transitioning from best friend to also being a colleague - hopefully, you'll have forgiven me for any monkey business. Marjory and Matthew, for a friendship that has endured the cold winter of Canada and the sunny shores of Nice... but eventually decided to move back to the cold winter part. Samuel Chéri, pour des soirées de rigolades inoubliables dans le froid du Québec (t'as vu, j'ai écris en français exprès pour toi).

My thanks are also to Salvatore, Mohit, Sandrine, Kevin and Ivo for the camaraderie of being a NeOpTer. To Anna for insider knowledge (all in good fun) about Neuroschool, and for her constant cheerfulness. To the upcoming generation of INT scientists who've set incredibly high standards : Cléo, Uriel, Miles, Marie. To the SYRT : Lucie, Montaine, Alexis and Lucio. To Taarabte, whose resourcefulness was nothing short of

a lifeline in saving my salary for the last year. To Hélène, Joelle, Céline, your collective kindness has been a shield against my administrative faux pas, and I am eternally grateful. To scientists outside my labs : Cyril Kurz, Ingrid Bureau, Melissa Mardelle, Emna Marouna, Jérémy Camon; Francesca Sargolini, Frank Vidal, Christian Gestreau – your influence has been crucial in setting me onto this path. To those who acted as lighthouses on this scientific voyage: the remarkable 2022 Schloss Rauischholzhausen crew (hello Crème de la Crème), the brilliant minds of the 2023 Telluride Neuromorphic Workshop (here's looking at you, Condo 301, the residence of champions), and the many other talented individuals I've had the good fortune of meeting at various scientific events. In that spirit, a final scientific salute goes to Tony Movshon, Dario Ringach, Peter Roelfsema, and the extraordinary Paolo Papale, for fruitful scientific exchanges that still carry impactful consequences on this day.

Finally, there remains the matter to acknowledge those who were not on the ship, but have offered unwavering support from the shore. To my dad, who merits the first position in this section of the acknowledgements, because the inspired 'Paperback Writer' quote was stolen from him, and also for pushing me to try this "science thing". Equally, to my mom, who is second in sequence but not in importance, who never faltered in her support to me and kept my cheese and happiness supply steady. I could have never done this without you ! To my girlfriend, who stood by me during the final push towards the finish line and tried to talk me out of my new cowboy hat-wearing habit since my return from Telluride. And then, of course, to my friends, ranked according to how much I appreciate them: Alexandre, Anthony, Arman, Axel, Cédric, Cem, Chloé, Dan, Davidou, Elysa, Florian, JB, Jérémie, Jessica, Louison, Manu, Margot, Martine, Michou, Nicolas, Perrine, Raphael, Rémi, Rich, Thomas, Tony, Victoria, Yannis and Zoé (I am of course kidding, that was in alphabetical order).

On a reflective note, and an important one, I want to express my gratitude to Umit Keysan for the memorable moments we shared. We playfully challenged each other to see who would earn their PhD first. While fate took an unexpected turn, the journey was enriched by your camaraderie. Your spirit continues to inspire me, and I miss you, my brother.

Last, but not least, thank you for reading this manuscript. Whether you are now reviewing a fresh copy before my defense, or are years later being forced to sift through every page to continue my project in one of the labs I was part of, thank you for breathing life once more into this work. I hope you will enjoy the reading as much as I enjoyed the *science-ing*.

Contents

List of Figures

List of acronyms

KL
　Kullback-Leibler. 41, 42, 44

LGN
　Lateral Geniculate Nucleus. 15, 25, 27–31, 57, 90, 91, 133, 134, 169

MUA
　Multi-Unit Activity. 117, 118, 123–125

V1
　Primary Visual Cortex. 15, 16, 18, 24, 25, 28–31, 33, 35, 45, 48–50, 52, 54, 55, 57, 62, 86, 88–93, 95, 112, 115, 117, 118, 123–128, 130–135, 137, 160, 161, 165, 167–170, 172, 224

1. Manuscript introduction

> *"The journey is the reward."*
> Daoist proverb, attributed to Laozi

1.1. Overview

The objective of this thesis is to establish a robust framework for probabilistic computations in vision, encompassing stochastic visual stimulation and advanced data analysis, to enhance our understanding of the neurophysiological mechanisms in V1 and beyond. We will center primarily on computations related to the variance of visual inputs, which play a fundamental role in our daily life, as highlighted in Chapter 2.

1.2. Scientific context

This work is based on a "cotutelle" (joint) supervision between Laurent Perrinet from the Institut de Neurosciences de la Timone laboratory (INT, Aix-Marseille University & CNRS), and Christian Casanova, initially based at the Laboratory of Visual Neurosciences (School of Optometry, University of Montreal), now also affiliated with the "École de Technologie Supérieure" (Montreal). This international and interdisciplinary endeavor combines French computational models with Canadian neurophysiological experiments. Over time, this project has also expanded to include several collaborators, such as Nicholas Priebe from the University of Texas at Austin, who conducted preliminary experiments in marmosets, as well as Pieter Roelfsema and Paolo Papale from the Netherlands Institute of Neuroscience, who performed similar recordings in awake behaving macaques.

The interplay between neurobiology and neurocomputations is the keystone of modern neuroscientific research. In line with this, this thesis aims to demonstrate the advantages of a model-driven approach to studying an intricate dynamical system such as the brain, advocating in the process for improved stimuli and generic models. While this manuscript should (hopefully) be accessible for readers with an interdisciplinary background, I have made every effort to ensure that the content is comprehensible regardless of whether the reader's background is more aligned with the silicon or carbon side of the field.

1.3. Manuscript organization

By virtue of the interdisciplinary and international organization of my thesis, this manuscript is longer than most manuscripts have a right to be. Despite this length, a concerted effort has been made to ensure a smooth and engaging reading experience. Rather than beginning with a lengthy introduction aimed at providing a broad overview of the nature of vision and its modern challenges, I instead elected to start with a concise Marr-like approach [200] to the problem at hand. After this introduction, each chapter is organized as a self-contained read, beginning with a concise introduction regarding the specific sub-problem studied, followed by a related article, and concluded with research perspectives.

This manuscript is structured into eight sections, as outlined below:

— In the current chapter, *Chapter 1*, the general study is introduced, with the aim of providing a high-level overview of the whole thesis in a comprehensive manner.

— *Chapter 2* starts the scientific content of the thesis, by introducing the notions on which further chapters are build. Given the multidisciplinary nature of said chapters, this global introduction aims at being a guide for the reader, providing introductory notions of the high-level content of this thesis. The specific details - or the specific experimental methodologies - of each such concepts are then explored in the introductory section of each of the following chapters.
By analogy with David Marr's levels of analysis [200], this introduction begins with the computational nature of perception and vision, then dives into the algorithmic nature of such a perceptual problem, to finally conclude with its implementation at a computational and neurobiological level.

— *Chapter 3* marks the first of our scientific contributions. It introduces the idea that the images that make up our daily visual experience, so-called "natural images", can be accurately described as a mixture of Gaussian-like distributions in orientation space. In the specific context of the thesis, this serves to support the use of artificial stimuli in the following chapter, in order to deal away with a number of complexity involved in the use of natural images as stimuli in visual experiments [263]. Building upon this concept, we also incorporate such variance in a computational model to demonstrate enhanced performance in orientation encoding and innovative deep learning applications. Specifically, this chapter explores the role of aleatoric (input-bound) and epistemic (model-bound) variance in the encoding of natural images, using a convolutional sparse coding algorithm. The results reveal that the integration of oriented features across multiple levels of epistemic variance significantly boosts the accuracy of sparse coding for natural images. Finally, this chapter also proposes that hierarchical visual processing can benefit from variance encoding, by training a deep convolutional neural network on sparse-coded natural images datasets, and demonstrating that variance-encoded sparse code is not only as effective as conventional images, but also provides a more robust representation of naturalistic images.

— *Chapter 4* extends on similar Gaussian priors, using a parametric generative

model called Motion Clouds [189]. These stimuli allow probing for the neural correlates of variance (in orientation space) using extracellular recordings in the primary visual cortex of anesthetized cats. The chapter quantifies the dynamics of sensory variance modulations of these neurons and reports two novel archetypal neuronal responses, one of which is resilient to changes in variance and co-encodes the sensory feature and its variance, thereby improving the population encoding of orientation. The existence of these variance-specific responses is then accounted for by a computational model of intracortical recurrent connectivity. This chapter's article proposes that local recurrent circuits process variance as a generic computation, advancing our understanding of how the brain handles naturalistic inputs. Arguably, this is the keystone of our contributions during this thesis, and the results obtained here serve as the basis of the remaining chapters.

— *Chapter 5* builds upon the foundational concepts established previously to develop a model of graphs with a topology that depends on input variance. This is an extension of the type of neuronal recordings presented in Chapter 4, but with a significant shift in the data source, as the recordings here were carried using high-density matrices implanted in awake, behaving macaques from Pieter Roelfsema's group. The fact that such results exists in two species of two different taxonomic genera confirms the relevance of the findings of Chapter 4 : recurrent interactions between neighboring V1 neurons are sufficient to explain the observed phenomena. Considering that the recordings presented in this chapter are preliminary and sourced from a single macaque, the content here is not developed into a full article and is notably shorter compared to other chapters.

— *Chapter 6* relates to the *multiscale* aspect of this thesis. By diving into models of the subcortical pathways based on previous anatomical [1] and functional studies [292] of the laboratory, we first reveal the presence of alpha oscillation-gated activity in pulvinar, a critical component of predictive variance computations. This work sheds light on the intricate dynamics of the pulvinar, and its role in regulating the flow of information between visual cortical areas. This is extended in a second contribution, which reviews the current literature of pulvinar as an information hub and describes the pulvinar as a regulator of (inverse) variance computations throughout the entire visual hierarchy.

— *Chapter 7* briefly summarizes all the findings of previous chapters, revisiting key contributions of this thesis and highlighting their significance under the framework of predictive coding. We then explore potential future direction for further research, considering how these articles might be built upon and extended. Reflections upon the modes of investigation, their limit and the potential alternative approaches are also laid down. The broader impact of this work is also discussed in this chapter. It considers how these findings might influence the field of neuroscience at large, rather than being limited to visual neuroscience. It also contemplates the potential implications for related fields, such as clinical research and machine learning.

— The *Appendices* section provides derivations of equations pivotal to the conceptual framework discussed in this thesis. While the reading of this section is indeed optional, the formulation of these derivatives offer valuable insights into the foundational mathematics underlying the central concepts of the free energy principle. Additionally, the appendices include supplementary scientific material, such as press releases related to this research.

2. General Scientific Introduction

"A major challenge of any scientific endeavor is not only to provide good answers to the questions we ask about our world, but to find good questions to ask in the first place."
Paul Cisek, Resynthesizing behavior through phylogenetic refinement, 2019

Summary

2.1. The problem under study

An introduction to a complex problem is arguably simplified by the use of a well-defined general theory. Alas, while neuroscience lacks something similar to physics' standard model [63], it has a rich history of integrative frameworks designed for a particular sub-fields of investigations. David Marr's three levels of analysis, originally developed to understand vision[200], is one such integrative framework, that has since then been successfully applied to numerous other problems[204, 241]. To understand the complex "problem of vision", Marr proposed three levels of approach: the computational level (examining the system's purpose), the algorithmic level (exploring its

procedural mechanism), and the implementational level (detailing practical aspects). Our present introduction adheres to this logical structure.

The general aim of this manuscript is to understand the purpose of variance-related computation in the brain (computational level), the logic behind such computations (algorithmic level), and the methods to enable both brains and machines to perform these computations (implementational level). As we shall see in the following sections, our approach will naturally lean towards Bayesian inference as a framework to articulate the problem(s) of vision. Bayesian schemes, like related probabilistic methods [7], are extremely valuable in making predictions about experimental outcomes, but are not designed to provide mechanistic explanations for those outcomes. In that sense, the descriptive models used in this thesis work synergistically well with Marr's approach [59] for one key reason: the *formal independence* of the three levels. Practically, this means that different algorithms can be used to solve the same computational problem, and that different hardware (or biological substrates) can implement the same algorithms. Therefore, computational level questions are *formally independent* from other levels, allowing them to be addressed without constraints from other levels. It then becomes that the relationship between the three levels is one of *realization*, that is, of incremental assembly. In the specific case of variance computations, this will allow us to start focusing on the "what" without being preoccupied with the "why", and then similarly with the "how", thus building this introduction incrementally from the descriptive to the explanatory.

2.1.1. The computational level: Vision in an uncertain world

Marr's first level of analysis, the computational level, seeks to describe the purpose of the system under study. Rather than deriving a working theory from the data, this allows us to approach the theory first, then work our way towards facts progressively. This section of the introduction is thus concerned with the idea of framing vision as an uncertain process.

However, as with all theses that relate to a particular sensory system, we cannot escape the required description of the system under scrutiny, which is here the neurobiology of visual processing. While this arguably pertains more to the implementational aspect of Marr's three levels, and thus should be discussed last, starting by introducing the notion of "orientation selectivity" in V1 considerably simplifies the remainder of the introduction. Overall, we aim to be rather synthetic here, quickly working our way from the retina towards V1.

We then dive properly into the computational nature of the inputs to the visual system, be they natural or synthetic, with a focus on the sources of uncertainty in vision. Finally, we discuss probabilistic accounts of vision and their (numerous) successes, which will serve as a justification of the algorithmic approach, described in the Section 2.1.2.

2.1.1.1. The neurobiology of vision *ante* V1

With a few niche exceptions, vision is a sense present ubiquitously throughout in the animal domain [180]. Vision is nothing short of a marvel of biology, a sense that starts by capturing massless elements of electromagnetic radiations, the photons, eventually converting them into a coherent representation of our environment. The general computational goal of vision is conserved throughout various species, but its implementation varies extraordinarily: sixty pair of eyes in scallops [238], polarized facets optimized for tracking mating partners in horseflies [136], hyperspectral cameras in mantis shrimps [312], several segregated nervous ganglia for visual processes in octopuses [345]... In the case of primates, our specific tweaking of vision is the dedication of almost a third of our cortex to processing visual inputs [154]. As fascinating as shrimp vision might be, we will here focus on a human-centric account of vision, namely describing of the primate visual system (Figure 2.1). However, given that chapters 4 and 7 deal with neural data acquired from anesthetized cats, we will finish this section with a brief comparison of the two species, as animal models used commonly in visual neuroscience.

Vision relies on a sophisticated bilateral sensing organ, the eye, which can be likened to an extraordinarily intricate adjustable-focus camera (Figure 2.2). Light enters the eye via the cornea, continues through the transparent aqueous humor, and proceeds through the pupil, which the opening of an aperture-regulating diaphragm, the iris. The focusing of the light source is then performed by the biconvex lens of the crystalline, which is dynamically adjusted by tiny muscles that modify the focal length. The nervous system thus possesses control over these components, enabling visual
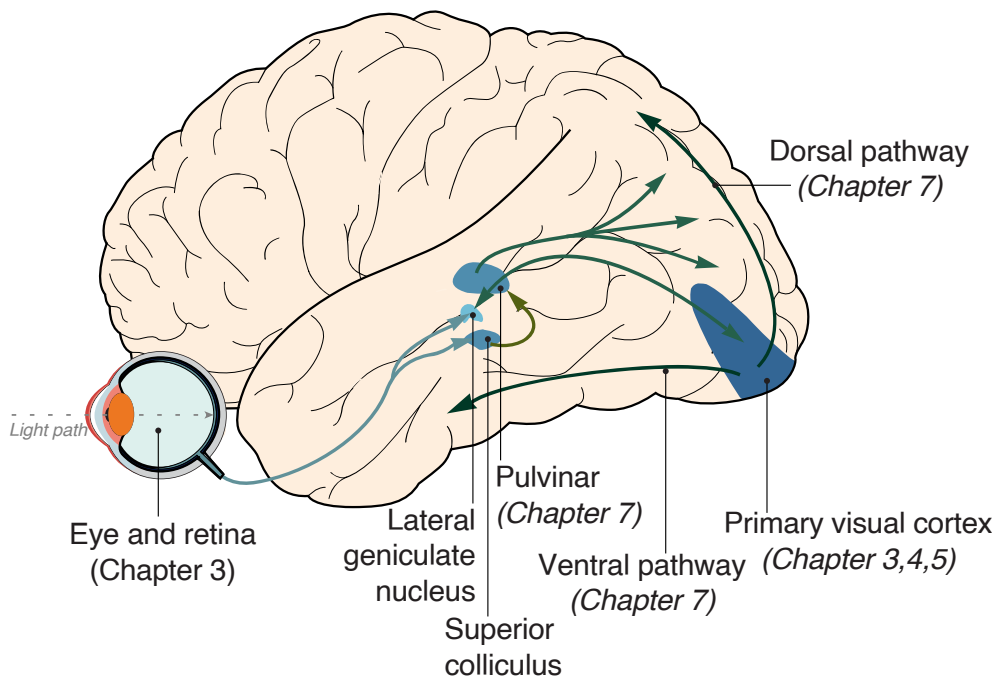
Figure 2.1. – A schematic overview of the visual system. Electromagnetic waves (photons) are transduced in the retina into binary neural activity (spikes), which are then transmitted through the optic nerve to the LGN, and then to V1, before being processed by further cortical areas. The present chapter introduces general notions to the visual system, with specific part discussed further in the chapters indicated here in *italics*. Figure adapted from [154].

signals to act as feedback for adjusting the eye's optical characteristics in a closed-loop system [84].
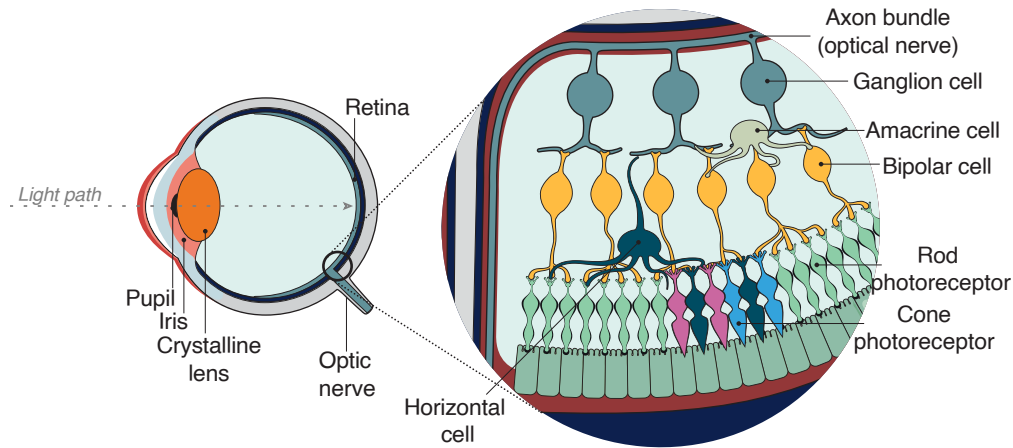
Figure 2.2. – Illustration of the eye and retina. Adapted from [133].

Only then does light reach the photosensitive surface situated at the rear of the eye: the retina. Interestingly, the light must also traverse the full depth of the retina before being actually captured, as the light-sensitive part of the retina lies in its depth rather than on its surface. This propagation through the retina disperses the light, resulting in a somewhat lower-quality image. However, this is compensated for by a particular type of cell that channels light [225], but also contributes to a space-efficient design [174]. A unique feature resulting from this architecture is that the axons from the retina's neurons must exit through the posterior section of the eye, and in doing so, must traverse the retinal depth in its entirety in some place. This is done in a single location, known as a "scotoma", which is thus devoid of photoreceptors and lacks direct visual input. However, this visual gap is seamlessly filled in, rendering it unnoticed in our conscious experience [259].

Aside to this particular design, the retina is a remarkably effective sensor. Specialized cells, the photoreceptors, convert light into electrical signals via a photosensitive molecule named retinal. These photoreceptors cells come in two varieties: cones, which cover different portions of the visible spectrum, providing important color visual information that we shall not be concerned with in this thesis [107]; and rods, which are useful in situations with low light intensity. These photoreceptors are distributed in a particular pattern throughout the retina, where the densest concentration of cones is found near the center of the visual field, in the "fovea". This means that in order to obtain the maximum density of light-sensing cells, and thus extract the maximum amount of visual information, the eye must hover in to points of interests [96], which occurs about every hundred of milliseconds in primates [141].

The signals from these two types of cells are conveyed in a two-stream parallel process: either to the bipolar cells, carrying feedforward signal processing [112], or to the horizontal cells, which carry out lateral integration in the retina, a process

that contributes to enhancements in contrast [202] and the early recognition of motion [233, 192]. The output of this whole early sensory processing is carried out by retinal ganglion cells, whose axons are bundled into the optic nerve and sent deeper into the brain.

This nerve eventually reaches the brain's central hub, the thalamus. In the particular case of the optic nerve, the dedicated thalamus nucleus that receives inputs from the optic nerve is the LGN, which is located in the posterior part of the thalamus. The wiring from the retina to the LGN is not straightforward, in the literal sense, as part of both eyes' field of view will cross each other over before reaching the thalamus, likely due to an evolutionary twist of the body plan that also resulted in similar decussation (crossing of the middle plane) in the spine [166, 182]. To simply label the thalamus as a relay would be a major oversimplification [279], as the inputs from the retina to the thalamus represent less than one tenth of the synapses [111], which might translate into even less than one tenth of actual functional impact [36].
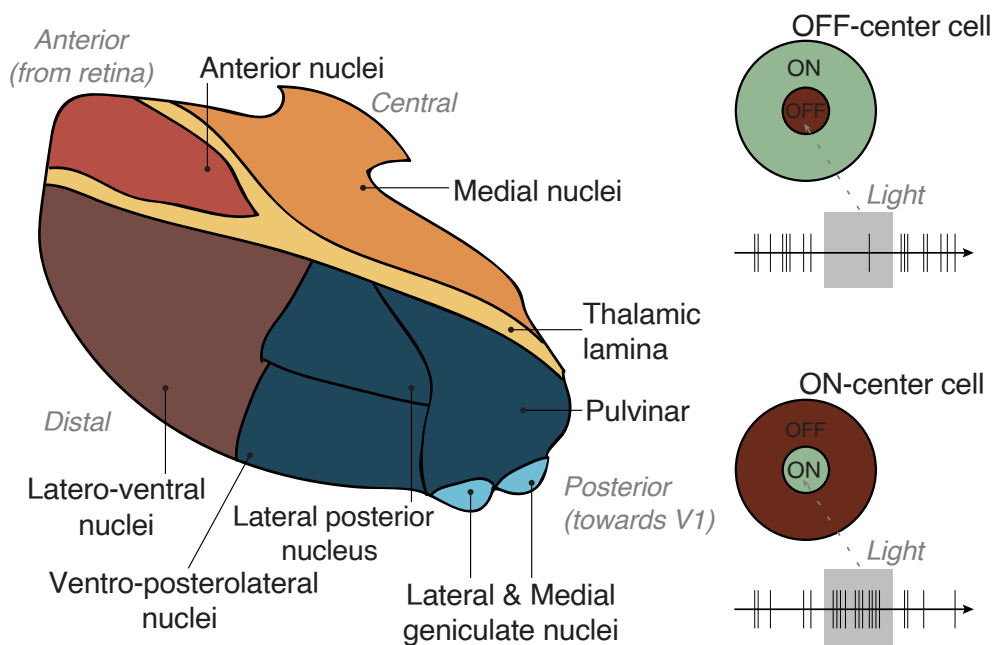
Figure 2.3. – (left) Anatomy of the LGN and nearby thalamic nuclei. Afferences from the optic nerve arrive at the LGN. The superior colliculus, located near the cerebellum, is not represented. (right) Illustration of the ON/OFF activation of neurons in the LGN.

Effectively, the LGN acts as the preliminary stage for information consolidation, characterized by cells that perform an ON/OFF transformation of visual input, inherited from the mode of action of retinal ganglion cells (Figure 2.3). ON-center cells are triggered by an increase in light intensity within the central portion of their responsive visual field, known as the "receptive field". Conversely, they are deactivated by an

increase in light intensity at the periphery of their receptive field. OFF-center cells function in the opposite manner. This shift signifies the transition from encoding unipolar light intensity in the early retina to bipolar light intensity in the thalamus, a crucial transformation for encoding contrast, which paves the way for encoding complex features in the cortex.

Throughout this thesis, this notion of "receptive fields" (mostly in V1) will be central to the subject of study. It is thus worth to define it clearly here, and for that, we will follow its textbook definition as the region in the sensory space in which a stimulus modulates the activity of a neuron. For Sir Charles Sherrington, who coined this term, a receptive field was exemplified by the skin area that, when stimulated, modulated the activity of scratch reflex neurons in canines [281]. In our context, a receptive field will always pertain to the spatial configuration of a visual stimulus that modulates the activity of a given neuron in V1.

Before we progress to our ultimate stop, V1, it's important to recognize the existence of what are referred to as "non-canonical" pathways from the retina. The superior colliculus, for example, constitutes about 10% of the visual output of the retina in primates, though this proportion is increased to 50% for the cat [234]. Similarly, the pulvinar nucleus of the thalamus, while only briefly mentioned here, warrants major discussion in the introduction to chapter 7, as it plays an integrative role among various cortical areas while also integrating inputs from the retina [46].

2.1.1.2. The neurobiology of vision *intra* V1

Following this five-page overview of precortical vision stages, we now shift our focus to the primary subject of this thesis: the primary visual cortex, V1. Located in the most posterior region of the brain, the occipital lobe, it is also known as the striate cortex, due to the presence of the line of Gennari [114], a massive bundle of axons originating from the LGN that forms a stripe that can be observed with the naked eye.

In terms of anatomy, like all non-motor cortices [284], V1 is composed of six layers. Despite ongoing debates surrounding the concept [135], the "canonical" circuitry of this layered cortex suggests that visual information enters from the LGN into the IVth cortical layer, is transported to layers II and III, and subsequently reaches layers V and VI [28, 73]. As we will see later in this introduction, this seemingly simplistic anatomical description is highly beneficial, as this pattern of connectivity is replicated across the whole cortex and can be thus mapped to general computations [74, 20]. Organized orthogonally with respect to the cortical surface, this pattern establishes what is often referred to as a "cortical microcolumn.". Depending on the anatomist and the definition, its diameter fluctuates between 200 and 800 μm [216]. Beyond this local anatomy, one should also note that layers II and III extend sizeable "lateral" axons, forming a "horizontal" circuitry [11, 51] that interconnects multiple similar microcolumns.

One can then naturally inquire about the functionality of these aforementioned circuit motifs. The main functional description of V1 can be traced back to a seminal study by Hubel and Wiesel, who conducted recordings from individual neurons in this
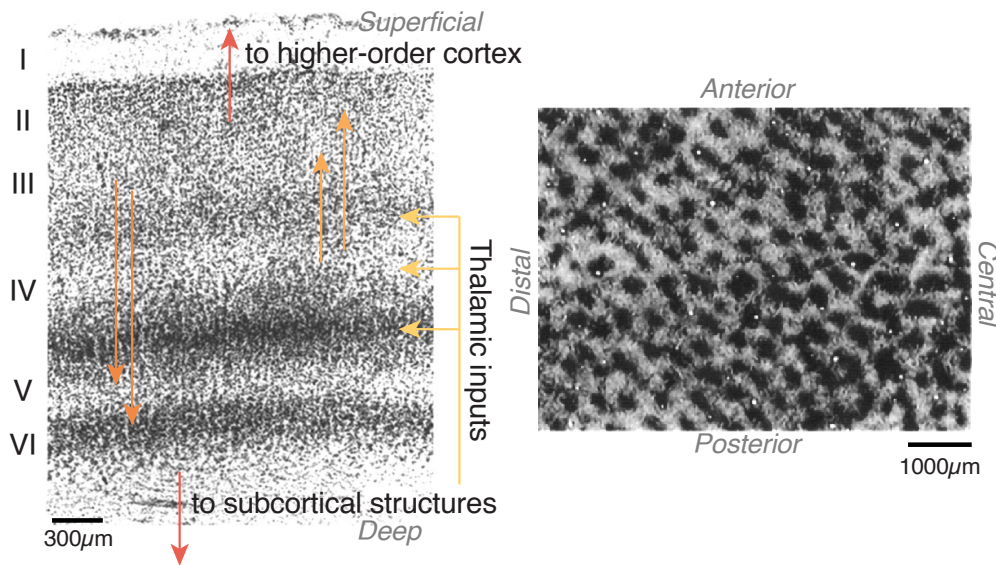
Figure 2.4. – Primary visual cortex anatomy. (left) Nissl staining showing the vertical organization of macaque V1, adapted from [271]. (right) Cytochrome oxydase staining, reflecting the horizontal organization of the thalamic inputs [310], adapted from [218].

area and discovered neural responses elicited for light bar of specific orientation [137, 138]. Functionally, this feature expands upon the ON/OFF structure found in the LGN, but incorporates a spatial layout that is lengthened, and thus appropriate for detecting edges in the visual field. Two distinct classes of cells demonstrate this "orientation selectivity": simple and complex cells. Simple cells, aptly named, react to edges of various orientations, and can be thought of as the convergence of many LGN cells. Complex cells, on the other hand, exhibit overlapping ON and OFF fields and some degree of spatial invariance, and can be conceptually understood as a hierarchical [138] (or lateral [48]) convergence of several simple cells [34]. Coherently, the majority of simple cells are found in layer IV and VI, whilst complex cells tend to be confined to layers II and III [240]. At the broader (mesoscale) level, the selectivity to orientation is arranged into maps of preferred orientation [120], where adjacent cells encode different orientations. This arrangement can be seen as essentially a competition for the best representation of objects that are present at a specific point within the visual field [159]. For a further detailed account of orientation selectivity, we refer the reader to chapter 4's introduction.

While orientation selectivity is the core concern of this thesis, there exists also other visual properties encoded by V1. A non-exhaustive list would contain the spatial frequency [318], referring to the scale of a specific feature and to its repeated size in the visual field. Like orientation, spatial frequency is also represented in a spatial frequency map that seems to be independent of orientation selective maps [143]. Possibly the most critical visual property encoded by V1 asides orientation is visual eccentricity
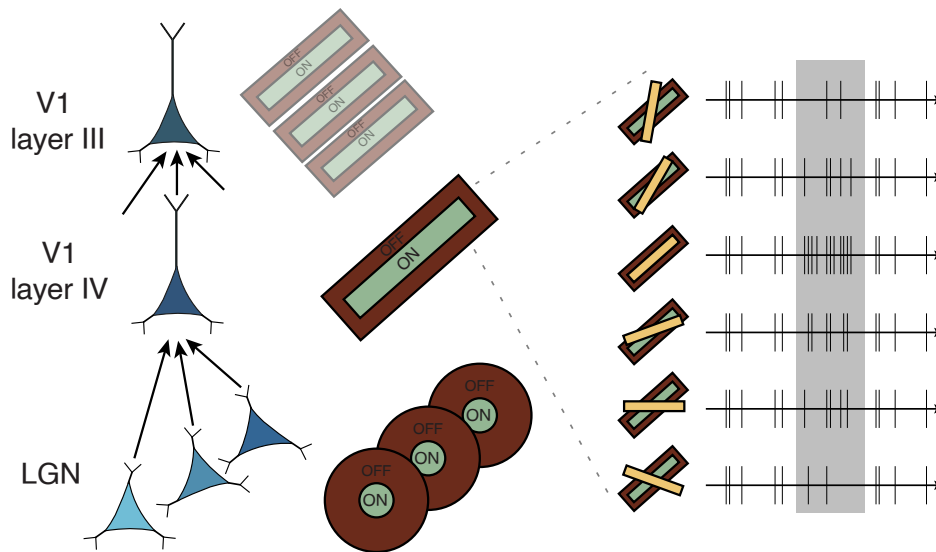
Figure 2.5. – Illustration of orientation selectivity in V1, as a hierarchical model [138] from unoriented LGN neurons, to simple then complex V1 cells.

and elevation, i.e. the spatial localization of visual stimuli [335]. In primates, there is a specific and highly ordered mapping of the visual field onto the topology of V1 that is well-captured by a mathematical "log-polar" transformation [319]. The closer a particular region of V1 is to the midline border, the more centrally located within the fovea that region's corresponding visual field will be. This precise arrangement allows for the efficient processing of visual information, enabling dense dedicated regions in the retina to have similar dense dedicated processing region in V1. This density spreads throughout the cortical area, thereby providing an "over-representation" of the central portion of the visual field (encoded by the foveal region), with respect to its peripheral portions [145].

V1 is not solely influenced by feedforward and local lateral interactions. It is, in fact, an integral component of a meticulously organized hierarchical network [267]. V1 extends projections to various "extrastriate" cortical areas, such as V2, V4, and IT, as well as to subcortical areas, such as the pulvinar. While the role of these pathways will be occasionally referenced throughout the articles compiled in this thesis, they are not a primary focus of ours, and will be thus reserved for their respective dedicated chapters (4 and 7).

As these chapters also deal with neural recordings gathered from anesthetized cats, one should note a few differences between theirs and the primates' visual systems. While both possess a primary visual cortex, or V1, primates have a greater number of color-detecting cones leading to better color perception, whereas cats have more rods [297, 171] aiding their low-light vision and 3D motion estimations, due to their nocturnal niche.

The main difference that concerns us however lies in V1. At the macroscale, the cat's

V1 is composed of Broadmann areas 17 and 18, instead of a single area in primate. Area 17, which was the targeted recording region in chapter 4, receives signals from the cat's equivalent of primate LGN's P cells (X in cats, concerned with fine details) and M cells (Y in cat, concerned with motion information) [301]. Area 18 on the other hand mainly receives axons from the Y cells [240]. At the mesoscale, likely to their respective cortical sizes and number of neurons [158], the cortical maps in V1 have different shapes [269], but both species show columnar organization, unlike rodents [151]. At the microscale, it seems that the general pattern of connectivity is similar, but with difference in specificities of laminar connections, namely from layer IV to III [195]. The reason for this seemingly mixed model usage in the literature is partly accounted by the fact that cats have been the main historical model under study. As such, the wealth of 60 years of data gathered in the cat visual system, combined with the complexity involved in primate experiments, still makes them highly relevant to vision research.

2.1.1.3. The statistics of natural and naturalistic visual inputs

Having introduced the basic notions of orientation selectivity and the neurobiological substrates of vision, we can now turn back to the goal we had set in the first place: provide a computational account of vision.

Understanding what vision does, without understanding what type of input vision must process, would be a daunting task. Vision is tasked with processing the complexity and richness of the images that form our everyday perceptual experiences, colloquially known as "natural images". These images, despite their immense diversity and complexity, are all governed by the same physical laws [300], leading to observable statistical regularities. This proves to be an essential feature, given that the neural activity required to process these images is energetically costly [184]. Further, any sensory processing is under the constraint of extreme time efficiency required for the survival of our organisms [315, 167]. Horace Barlow [18] posited that up to V1, the purpose of vision is to eliminate such statistically redundant data, thereby optimizing energy usage and speeding up computations. This perspective, often referred to as efficient coding [231] allows providing a quantitative description of sensory systems with respect to their environments [286].

We will here overview such quantitative accounts to get a better sense of the structure of the visual world. Such structure will eventually lead us to a formulation of vision as a variance-bound problem. Fundamentally, vision concerns itself with the perception of light patterns and is, therefore, inherently constrained by the statistical characteristics of light intensity. Most, but not all, (namely in terms of contrast) of these light intensity statistics are primarily the problem of the retina [17]. Even though they are not the focal point of this thesis, it is crucial to acknowledge them, as the statistical regularities of vision are hierarchical, and thus the ones of light patterns provide substantial insight into the higher-order statistics of vision. In terms of pure lighting intensity, the distribution of natural images follows a Gaussian structure [183]. In line with Barlow's efficient coding hypothesis [18], the cumulative probabilities of such distribution actually mirrors the response to contrast in the retina's [223, 183].
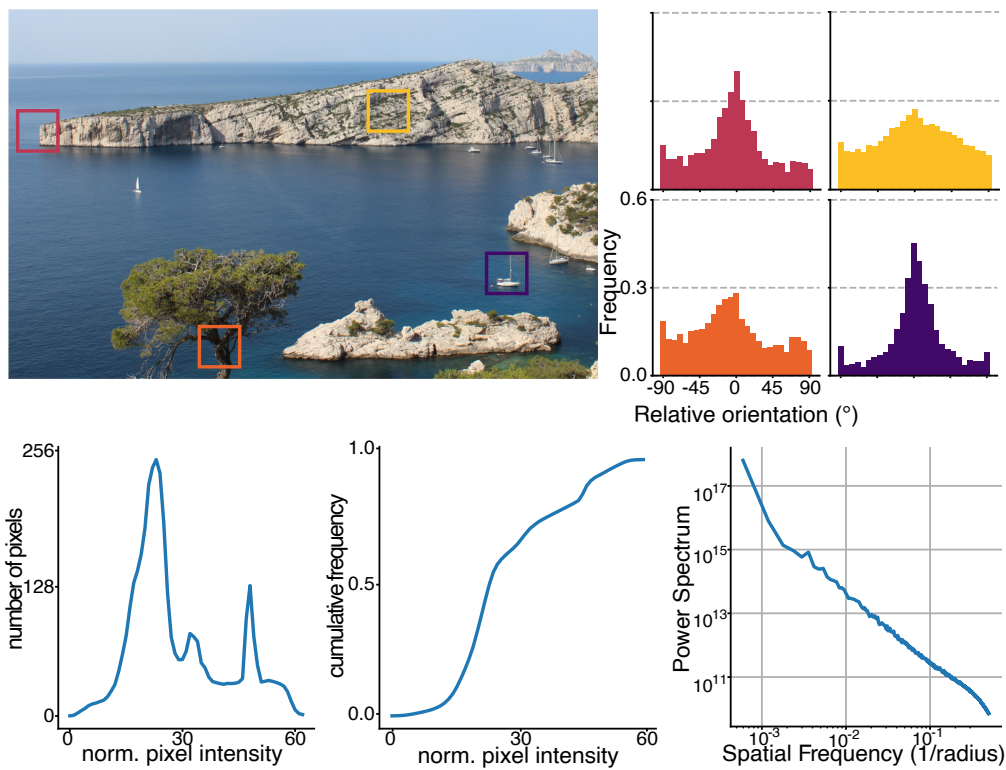
Figure 2.6. – (top) An example of natural image, Marseille's "calanques" [178], with local distribution of orientations shown on the right for multiple patches. (bottom) The statistics of this image, from left to right: distribution of luminance, cumulative luminance frequency, $1/f^2$ Fourier power spectrum.

This suggests an optimal tuning of the retina to respond to the intensity of light, underlying the interplay between the statistical properties of the natural environment and the sensory systems evolved to process them. This same kind of input/system "echo" will be further described in chapter 3.

The spatial organization of light intensity presents further statistical regularities that can help our posing of the vision problem. If one is to read this manuscript in a laboratory, an intuitive observation that can be had from gazing just about anywhere in a modern monochrome office space is the strong correlation in light intensity between neighboring locations. The same is also true, to a lesser extent, in natural images. Applying a Fourier transform to analyze such an image in the frequency domain, the correlation in intensity between nearby pixels becomes apparent in the form of a concentration of power in the lower frequencies of the spectrum. These low frequencies correspond to larger, more global structures, while the less powerful higher frequencies denote smaller, localized variations. This relationship is captured by a $1/f^x$ relationship characteristic observed in natural images, where often $x \approx 2$ [317].

This leads us to the notion that natural images also contain useful statistical regularities for higher-order features, like orientation. A noteworthy empirical demonstration by Bruno Olshausen and David Field [232, 230] reveals that when trained on natural images, reconstructive models will naturally yield oriented filters that are strikingly (and quantitatively) similar to those of V1. The statistics of such edges even enables prediction of high-level features in an efficient and unsupervised manner [245], possibly priming many downstream computations. In orientation space, this representation is highly sparse, with only a selected few orientations required at each spatial localization to reconstruct an image [87]. In feature space, however, the distribution of oriented features in natural images remain underexplored in existing literature, with a few exceptions. Mathematically, a single image embodies infinite uncertainty, with an uncharacterized distribution featuring high activations around the horizontal and vertical axes [124]. In chapter 3, we propose a more detailed exploration of the statistical structure of these oriented features, with the prevailing assumption being that the distribution follows a Gaussian pattern in orientation space.

This final note gives us an opportunity to keep statistical regularities in mind when talking about vision, whilst also getting essentially rid of natural images. While we briefly detailed their useful statistical regularities here, they are mostly just poised with intractable complexity [106, 263]. Because of this, and because of historical technical complexities, vision has mainly experimented with artificial stimuli since its inception 60 years ago. This has given investigations a methodology that affords a degree of parametric control over the multifaceted system of vision [249], but has had the unfortunate drawback of not eliciting the same types of activation as natural images [90, 344, 23]. This enduring debate between the use of "natural" versus "artificial" stimuli leads us to a using a satisfying compromise in this thesis: the "naturalistic" stimuli. Such experimental approach aims to strike a balance between replicating the statistical properties of natural scenes and maintaining the experimental control offered by fully artificial stimuli, thereby providing an optimal platform for investigating the intricacies of vision.
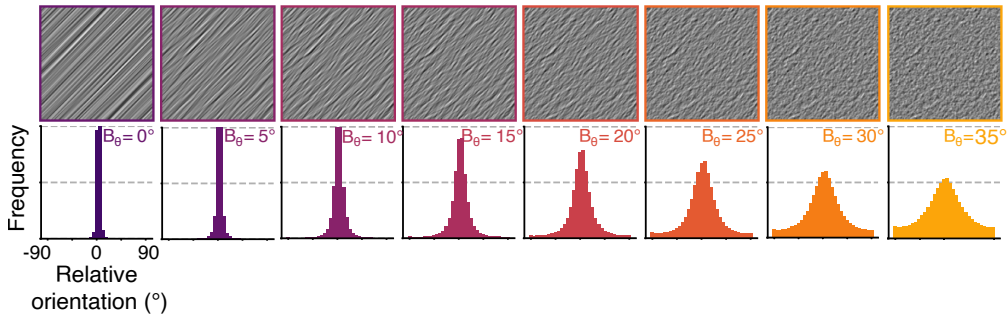
Figure 2.7. – An example of Motion Clouds, with increasing orientation variance from left to right, and associated orientation distributions.

Here, we will focus on the case of Motion Clouds, generative model-based stimuli [189] which allow for fine parameterized control over a naturalistic stimuli [322], which is a desirable trait when probing sensory systems under realistic conditions [263]. They are mathematically defined as band-pass filtered white noise stimuli, whose filters in Fourier space are defined as a parameterized distribution in a given perceptual axis (here, only orientation). Thus, the Motion Clouds presently used are fully characterized by their mean orientation and their orientation variance, such that a given stimulus S can be defined as:

$$S = \mathscr{F}^{-1}(O(\theta, B_\theta)) \tag{2.1}$$

where $\mathscr{F}$ is the Fourier transform and O the orientation envelope, characterized by its mean orientation θ and its orientation bandwidth B_θ. For $B_\theta < 45.0°$, a good approximation is $B_\theta = 1/\sqrt{\kappa}$, where κ is the concentration parameter of a von Mises distribution (see below), and hence approximates the standard deviation [308]. It thus serves as a measure of the orientation variability in the pattern, and as such, we will use the term "variance" to describe it throughout the thesis. The orientation envelope is a von Mises distribution:

$$O(\theta, B_\theta) = \exp\left\{\frac{\cos\bigl(2(\theta_f - \theta)\bigr)}{4 \cdot B_\theta^2}\right\} \tag{2.2}$$

where θ_f is the angle of the frequency components of the envelope in the Fourier plane, which controls the spatial frequency parameters of the stimuli. For the range of values of B_θ considered in the present thesis, the orientation envelope approximates a Gaussian distribution and B_θ is thus a measure of the variance of the orientation content of the stimuli that follows a naturalistic distribution, as highlighted in chapters 3, 4 and 5.

Such stimuli offer three advantages over both simple grating-like stimuli and complex natural images. First, they enable fine control of mean orientation, controlled by θ, and its variance, controlled by B_θ. This thus allows to reproduce natural images' oriented content, solely in terms of orientation distributions. Second, as they are

stationary in the spatial domain, they only probe orientation space, excluding any second-order information exploitable by the visual cortex [147]. Third, by conforming to natural images' $1/f^2$ power spectrum distribution [87], they attain a desirable balance between controllability and naturalness [263]. The combination of these traits allowed Motion Clouds to be successful in describing the perceptual integration of the human visual system [287], the dynamical computations of retina [255], and the dynamics of naturalistic perception in V1 [321]. Controlling a naturalistic stimulus confers a significant benefit here, namely in allowing us to control the variance of these distributions, whose role in vision we will now explore.

2.1.1.4. Variance in vision

Herman von Helmholtz, in his treatise of physiological optics [130], gave a pinpoint accurate description of variance in sensory perception:

> *"Given that the world we live in is loaded with statistical noise, [≈ variance] expectations must be represented as part of the brain's models."*

Sources of variance are plentiful in vision (Figure 2.8). Well-known optical illusions demonstrate how an object can seem brighter than it actually is, when light intensity is juxtaposed against a dark background. Obstruction patterns lead us to inaccurately fill in parts of the visual field. Motion can blur our perception. Here, our primary interest lies in the orientation variance and its effect on vision, a sub-field that has been extensively explored psychophysically, yet remains relatively uncharted in terms of electrophysiological investigations.
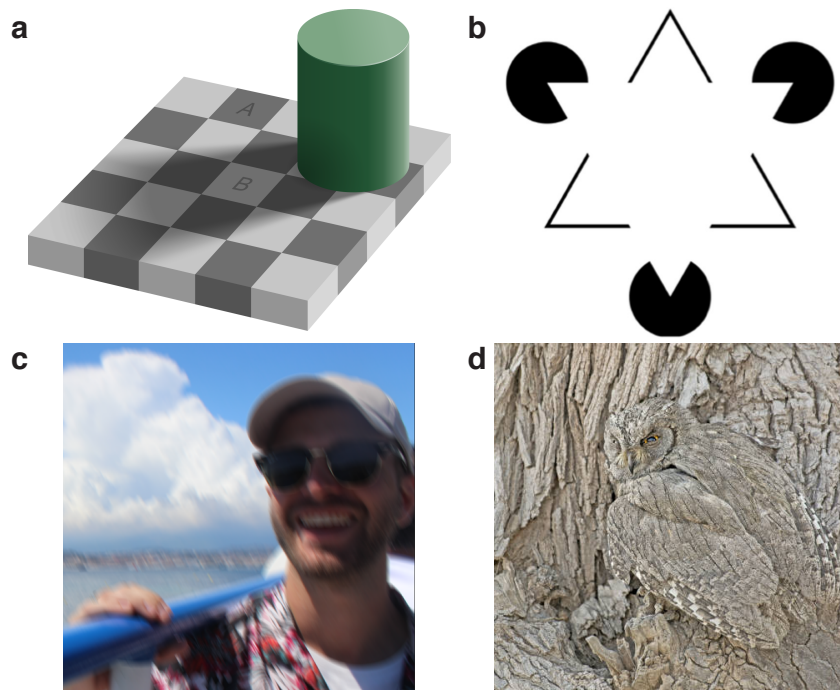
Figure 2.8. – Example of visual uncertainties, on: (a) contrast, (b) obstruction, (c) motion blur, (d) orientation (picture taken by Noor Hussain).

In psychophysics, variance is often interchangeably used with the term 'uncertainty' [309]. In the context of orientation, the term 'bandwidth' is also often used, for historical reason [291], as studying a broadband signal in orientation domain is straightforward, and creating such a signal by blending point sources of orientation to generate a distribution is even simpler. In statistical term, one can also use inverse variance ($1/\sigma$) and its squared term, the precision ($1/\sigma^2$). Under Gaussian approximation, such as in chapter 3 and 4 (see also Equation 2.2), bandwidth approximates variance, and both can be linked to inverse variance and precision through their straightforward mathematical relationship. Uncertainty, when used in this manuscript, will also refer to the notion of a "broader" distribution of features (often, of orientations). One could nonetheless be "certain" of a given large "variance", for example, by asking a human subject to judge the degree of variance of a given texture. That specific meaning will not be used in this manuscript (see for example [19, 109, 266]), and we will refer to inverse variance and variance when describing the computations at play here.

The general study of orientation variance can be traced back to parametric texture generation, starting with Bela Julesz work on "textons" [149]. These simple oriented elements were shown to human subjects, and shown to take an increasingly great amount of time to process as their complexity (i.e. variance in orientation space) increases [150]. Under different texture synthesis algorithms and experimental paradigms, this observation has been replicated a number of time, showing that the impact of increasing variance on orientation detection is rather intuitive: as variance rises, performance deteriorates [247, 128, 127]. This is a non-linear effect [127] that works similarly to contrast response curve. Computationally, it is well accounted for by models of recurrent cortical activity [161, 24], and more specifically the result from the competition among multiple orientation detectors [257].

This is however not a mere passive influence, but an active mechanism, accounted for by human observers [19]. This behavior aligns with Bayesian models of perception, which, as we will discuss in the upcoming algorithmic section of the introduction, is of central importance for our work. Briefly, a Bayesian account of perception implies an active encoding of variance in the brain, granting access to this information for decision-making [110]. For instance, let us imagine an observer walking through a forest and hearing leaves rustling. If one's vision is clear, and one spots a squirrel, they'd likely carry on without concern. However, if their vision is obstructed by dense bushes (creating an increased orientation variance), they'd likely rely more heavily on their internal priors, which in most instances, would dictate a cautious retreat. Vision does not work as passive acceptance of uncertain input, but rather, requires our brain to cross-reference what it sees with a prior model of the world.

Despite the convergence psychophysical accounts and theoretical requirements, there is a significant gap in the literature on the implementation of variance computations in the brain [169]. This discrepancy could be attributed to the complexities involved in applying parametric texture synthesis to probe the visual system, making electrophysiological investigations challenging. As this section of the introduction primarily focuses on the computational level, we direct the reader to the concluding part of the introduction for a detailed explanation of the implementational mecha-

nisms of variance within the brain, and now turn our attention to describing the basic requirements for variance-based in computations.

2.1.2. The algorithmic level: A probabilistic model of perception

Marr's second level, the algorithmic level, is concerned with the operations that the system performs, and how these are organized to realize the computations described at the previous level. This section aims to describe probabilistic operations that can account for visual tasks, adopting an iterative approach to the problem of perception. We will follow the common formalism and mathematical development for this class of problem [31], framed with additional justifications of equations and examples fitted to the problem under study.

2.1.2.1. A simple example of Bayesian Inference

We start by considering a simplistic toy example in which an organism is tasked with inferring a single variable, the size of a food item denoted by v, based on a single observation, the light intensity denoted by u. The organism's sensory input is constrained, as ours, to be noisy [17], such that the true item size v is normally distributed based on observations u. Given that light reflections are related to the surface of the object, let's consider a distribution with a mean true item size $g(v)$, where $g(v) = v^2$, and a variance based on noise on observations Σ_u. This is expressed as:

$$p(u|v) = f(u; g(v), \Sigma_u) = \frac{1}{\sqrt{2\pi\Sigma_u}} exp \frac{(u - g(v))^2}{2\Sigma_u} \tag{2.3}$$

Given this probabilistic description of the input and the arguments previously stated in favor of vision as a probabilistic process, we shall also describe such organism as performing probabilistic inference. We thus consider that this organism has had the chance of encountering many food items in his life, and that it can implement prior expectation $p(v)$ about the size of the food item. This is also normally distributed, with mean v_p and variance Σ_p:

$$p(v) = f(v; v_p, \Sigma_p) = \frac{1}{\sqrt{2\pi\Sigma_v}} \exp\left\{ \frac{(v - v_p)^2}{2\Sigma_p} \right\} \tag{2.4}$$

Given this framework, we can apply Bayes' theorem to compute an exact solution to the inference problem based on a single observation of light intensity, which is given by:

$$p(v|u) = \frac{p(v)p(u|v)}{p(u)} = \frac{\frac{1}{\sqrt{2\pi\Sigma_p}} \exp\left\{ \frac{(v-v_p)^2}{2\Sigma_p} \right\} \frac{1}{\sqrt{2\pi\Sigma_u}} \exp\left\{ \frac{(u-g(v))^2}{2\Sigma_u} \right\}}{\int p(v)p(u|v)dv} \tag{2.5}$$

The numerator term is simply the product of the prior and the likelihood described above. The normalization term, however, is more complex. It is the integral of all $p(v)p(u|v)$, which ensures that the posterior probabilities of $p(v|u)$ sum to 1, but is too computationally intensive for a realistic system, as it would require sampling all

possible configurations of the problem. This consideration is left to the implementational level of the introduction, and does not deter us presently from computing an exact solution using Bayes' theorem. This yields the graphical representation of integrating likelihood $p(u|v)$ and prior $p(v)$ into a posterior distribution $p(v|u)$.
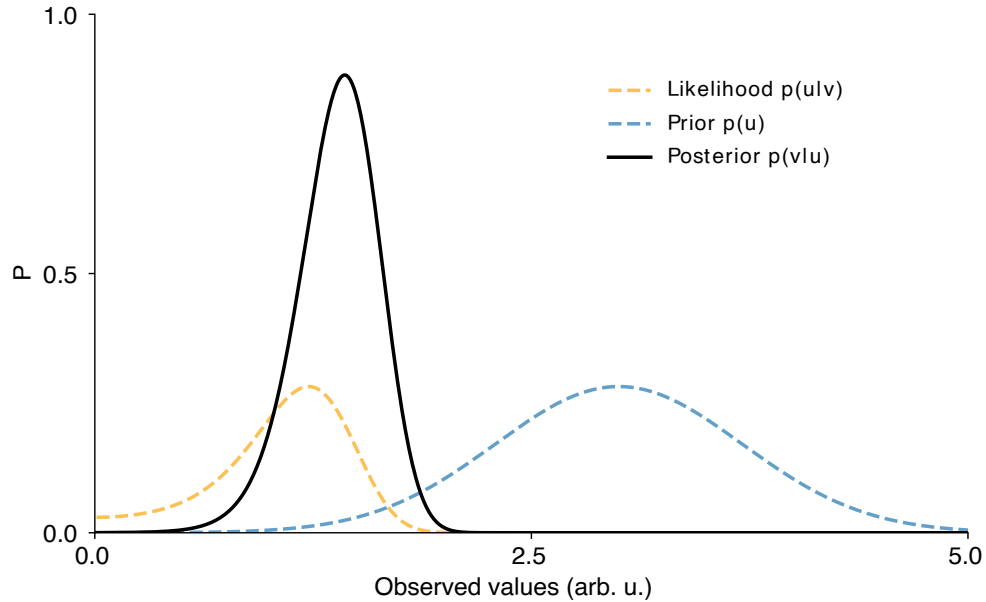
Figure 2.9. – A simple example of Bayesian integration. Integrating observed (likelihood) and former (prior) distributions, with $\Sigma_u = \Sigma_p = 0.5$. The most likely value of v lies between the two distributions.

Through the representation of Figure 2.9, one can notice how non-intuitive and non-linear Bayesian inference can be, even in such a simple example framed here. For the organism that must estimate $p(v|u)$, once the (input) likelihood $p(u|v)$ involved becomes a non-standard distribution, it becomes necessary to represent infinitely many $p(v|u)$ values for many possibles v, rather than first order statistics like mean and variance. As we've said, however, this would require infinitely too many computations for a biological system. Instead of computing the whole distribution, this organism could try to maximize the value of $p(v|u)$, in order to infer the most likely value of v. This procedure is called *maximum likelihood estimation* and involves sampling a posterior distribution $p(\Phi|u)$ of the most likely size Φ:

$$p(\phi|u) = \frac{p(\phi)\,p(u|\phi)}{p(u)} \tag{2.6}$$

As the denominator does not depend on Φ, we focus on the numerator. We will denote its logarithm F, as it relates to the Free Energy described later:

$$F = \ln\big(p(\Phi)p(u|\Phi)\big) = \ln p(\Phi) + \ln p(u|\Phi) \tag{2.7}$$

The natural logarithm having removed the exponential terms involved in Equation 2.5, we get:

$$F = \ln p(\Phi) + \ln p(u|\Phi)$$

$$= \ln\left(\frac{1}{\sqrt{2\pi\Sigma_p}}\exp\left\{-\frac{(\Phi - v_p)^2}{2\Sigma_p}\right\}\right) + \ln\left(\frac{1}{\sqrt{2\pi\Sigma_u}}\exp\left\{-\frac{(u - g(\Phi))^2}{2\Sigma_u}\right\}\right)$$

$$= \ln\left(\frac{1}{\sqrt{2\pi\Sigma_p}}\right) + \ln\left(\exp\left\{-\frac{(\Phi - v_p)^2}{2\Sigma_p}\right\}\right) + \ln\left(\frac{1}{\sqrt{2\pi\Sigma_u}}\right) + \ln\left(\exp\left\{-\frac{(u - g(\Phi))^2}{2\Sigma_u}\right\}\right)$$

$$= \ln\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}\ln\Sigma_p - \frac{(\Phi - v_p)^2}{2\Sigma_p} + \ln\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}\ln\Sigma_u - \frac{(u - g(\Phi))^2}{2\Sigma_u}$$

$$= \frac{1}{2}\left(-\ln\Sigma_p - \frac{(\Phi - v_p)^2}{\Sigma_p} - \ln\Sigma_u - \frac{(u - g(\Phi))^2}{\Sigma_u}\right) + C$$

$$\tag{2.8}$$

Now, it is simpler to find the derivative of F to look for a maximum value of our food estimate posterior distribution, we can compute the derivative of F over Φ (as done in Appendix A), we get:

$$\frac{\delta F}{\delta \Phi} = \frac{(u - g(\Phi))}{\Sigma_u}g'(\Phi) + \frac{(v_p - \Phi)}{\Sigma_p} \tag{2.9}$$

which endows our example organism with the ability to find the most likely value of the food item size v by varying Φ with respect to the sign of the derivative. The first term of the equation drives the posterior distribution towards the likelihood, and the second one towards the prior, both of which are weighted by the inverse of their respective variance, Σ. This equation is central to the remainder of this algorithmic part of the introduction, and really to this entire thesis, as it provides a mathematical role for variance in (visual) perception. It can be easily tied, in this form, to the Free Energy principle, introduced in the next section.

2.1.2.2. Free Energy principle for variational inference

The free energy principle allows both to understand how complex systems can model the most likely values of variables, as above, but also their distribution, as is the goal of this thesis. Briefly, this theory traces its roots to the work of Hermann von Helmhotz' unconscious inference [130], who was arguably the first physiologist to propose the notion that the mind construct a perception of the world through probabilistic inference. The idea of free energy in the brain gained popularity through the seminal works of Karl Friston [99, 157, 5], who also proposed its realization in the form of a neuroscience theory called predictive coding [92, 98]. In this section, we shall dive into the details of variational free energy, a specific formulation of free energy that will allow us to formulate general equations on which the rest of the thesis will rely.

In the text preceding Equation 2.6, we mentioned that posterior distributions $p(v|u)$

can have complex shapes that mandate prohibitively dense sampling for many values of u. As an approximation to avoid this issue, we proposed the method of sampling solely the maximum of these distributions. However, it is beneficial to know the shape of the full distribution, namely for knowing the variance - and thus the reliability - of estimates [197], as we wish to see in neurobiological systems in this thesis. An improved solution of maximum likelihood estimate is an alternative approach that involves the utilization of a "surrogate" distribution, in a process called variational inference [219]. This distribution, which we will refer to as $q(v)$, possesses a standard form that can be succinctly described by its mean and variance, unlike the distribution it surrogates in the first place. To approximate our complex posterior $p(v|u)$ with $q(v)$, we need a metric of (dis)similarity between the two distributions. For probability distribution and without particular assumptions, the choice is naturally the Kullback-Leibler (KL) divergence, defined as:

$$KL(q(v), p(v|u)) = \int q(v) \ln \frac{q(v)}{p(v|u)} dv \tag{2.10}$$

This divergence, or distance, is not symmetric, meaning that the divergence from $p(v|u)$ to $q(v)$ is not the same as the divergence from $q(v)$ to $p(v|u)$. This characteristic makes it particularly suitable for our purpose, as it allows us to measure how much information is lost when we use the surrogate function $q(v)$ to approximate the original distribution $p(v|u)$. By minimizing the KL divergence, we can ensure that our surrogate function is as close as possible to the original distribution, thereby providing a more accurate and efficient representation of the complex posterior distribution.

As is, this approach doesn't make things any easier, as we are still required to calculate the same normalization term as in Equation 2.5, a barrier that led us to maximum likelihood estimate in the first place. This is where the concept of Free Energy proves to be immensely beneficial. By definition, we already know that:

$$p(v|u) = \frac{p(u,v)}{p(u)} \tag{2.11}$$

which we can use into the equation of the KL divergence as:

$$\begin{aligned} KL(q(v), p(v|u)) &= \int q(v) \ln \frac{q(v)}{p(v|u)} dv \\ &= \int q(v) \ln \frac{q(v) p(u)}{p(u,v)} dv \\ &= \int q(v) \ln \frac{q(v)}{p(u,v)} dv + \int q(v) \ln p(u) dv \end{aligned} \tag{2.12}$$

given that $q(v)$ is a probability distribution and sums up to 1, we thus obtain:

$$KL(q(v), p(v|u)) = \int q(v) \ln \frac{q(v)}{p(u,v)} dv + \ln p(u) \tag{2.13}$$

by defining the variational free energy as the negative of the term that is concerned

with our surrogate distribution, we get:

$$F = \int q(v) \ln \frac{p(u,v)}{q(v)} \, dv,$$

$$KL(q(v), p(v|u)) = -F + \ln p(u)$$

(2.14)

since only F pertains to our surrogate distribution $q(v)$, the parameters of this function that minimize the divergence between $q(v)$ and the actual posterior $p(v|u)$ are the same as those that maximize F. Note that this F denotes that variational free energy, that is, the free energy involved in this approximation process, and not the general free energy of the system. To obtain the best surrogate distribution to approximate the posterior, we can now simply maximize $-F$, which does not involve the complex computation of the normalization term. In the next section, this will also serve to introduce learning of parameters of models. As we shall soon see, such learning involves, intuitively, being wrong as rarely as possible in our inference process. In other term, we wish to minimize the surprise, as defined by Shannon [277], associated with our predictions. Given Equation 2.14:

$$-\ln p(u) = F + KL(q(v), p(v|u))$$

(2.15)

As the KL divergence is positive, F can only be a lower bound of $\ln p(u)$. Maximizing the variational free energy F thus minimizes surprise $\ln p(u)$, meaning that we improve the approximation of $q(v)$ and thus optimize our internal model using this simple framework.

2.1.2.3. Prediction errors under the free energy principle

As we are progressively moving from the algorithmic to the implementation level, our formulation of Bayesian inference under the Free Energy principle could use a reframing more related to neurobiology. In that sense, this section of introduction will tell us now how to formulate Bayesian inference in simpler terms, and second how to derive learning rules.

Going back to Equation 2.9, we showed that deriving the parameter Φ could be written as:

$$\frac{\delta F}{\delta \Phi} = \frac{(u - g(\Phi))}{\Sigma_u} g'(\Phi) + \frac{(v_p - \Phi)}{\Sigma_p}$$

(2.16)

we can now perform some helpful variable renaming, with the left-hand side of the equation becoming:

$$\epsilon_p = \frac{(v_p - \Phi)}{\Sigma_p}$$

(2.17)

and the rest being:

$$\epsilon_u = \frac{(u - g(\Phi))}{\Sigma_u}$$

(2.18)

such that the derivative becomes:

$$\frac{\delta F}{\delta \Phi} = \epsilon_u g'(\Phi) + \epsilon_p \tag{2.19}$$

and the update rule of Φ to follow the gradient of F is then:

$$\dot{\Phi} = \epsilon_u g'(\Phi) - \epsilon_p \tag{2.20}$$

Both ϵ_p and ϵ_u measure the difference between a real and inferred value, and are thus both called prediction errors [92] (hence the term predictive coding). The term ϵ_p is often referred to, in the literature [211], as the prediction error on the causes, and measures the difference between the inferred observation and the model's prior expectations. The term ϵ_u, on the other hand, is called the prediction error on the states, and measures the difference between the observed value and the inferred value. In predictive coding, the goal of a model is to minimize both sources of prediction errors, and as such it needs a learning rule based on prediction errors. When looking for the point at which the system's prediction error become null, we get the stable point of ϵ_p:

$$\begin{aligned}\epsilon_p &= \frac{\Phi - v_p}{\Sigma_p} \\ \Sigma_p \epsilon_p &= \Phi - v_p \\ \Phi - v_p - \Sigma_p \epsilon_p &= 0\end{aligned} \tag{2.21}$$

and the one of ϵ_u:

$$\begin{aligned}\epsilon_u &= \frac{u - g(\Phi)}{\Sigma_u} \\ \Sigma_u \epsilon_u &= u - g(\Phi) \\ u - g(\phi) - \Sigma_u \varepsilon_u &= 0\end{aligned} \tag{2.22}$$

And thus the dynamics of a system that seeks to minimize its prediction errors can be described as:

$$\begin{aligned}\dot{\varepsilon_p} &= \phi - v_p - \Sigma_p \varepsilon_p \\ \dot{\varepsilon_u} &= u - g(\phi) - \Sigma_u \varepsilon_u\end{aligned} \tag{2.23}$$

This is however not an ideal description, because it assumes the rest of the parameters of the system v_p, Σ_p, Σ_u are constrained. It becomes even more inconvenient in a thesis about the implementation of dynamical computations on Σ parameters. This is where the free energy principle becomes extremely useful, as it ascribes a goal to the model: modifying parameters such that visual input u becomes the least surprising. As described in Equation 2.5, this is a term that involves the computation of a complex integral, but its natural logarithm can be more easily computed as in Equation 2.15, and even more so with F defined in Equation 2.8:

$$\ln p(u) = F + KL(q(v), p(v|u))$$

$$= \frac{1}{2} \left[-\ln \Sigma_p - \frac{(\Phi - v_p)^2}{\Sigma_p} - \ln \Sigma_u - \frac{(u - g(\Phi))^2}{\Sigma_u} \right] + C + KL(q(v), p(v|u)) \qquad (2.24)$$

as the KL divergence is a strictly positive term that forms a lower bound on surprise, we will include it in the constant C term for simplicity's sake. Following the (lenghty) derivations in Appendix A of F over v_p, Σ_p, Σ_u, we get the following dynamics of our inference model:

$$\frac{\delta F}{\delta v_p} = \frac{\Phi - v_p}{\Sigma_p}$$

$$\frac{\delta F}{\delta \Sigma_p} = \frac{1}{2} \left[\frac{(\Phi - v_p)^2}{\Sigma_p^2} - \frac{1}{\Sigma_p} \right] \qquad (2.25)$$

$$\frac{\delta F}{\delta \Sigma_u} = \frac{1}{2} \left[\frac{(u - g(\Phi))^2}{\Sigma_u^2} - \frac{1}{\Sigma_u} \right]$$

by re-expressing these equations using the definition of prediction error as before, we get:

$$\frac{\delta F}{\delta v_p} = \frac{\Phi - v_p}{\Sigma_p} = \epsilon_p$$

$$\frac{\delta F}{\delta \Sigma_p} = \frac{1}{2} \left[\frac{(\Phi - v_p)^2}{\Sigma_p^2} - \frac{1}{\Sigma_p} \right] = \frac{1}{2} \left(\epsilon_p^2 - \Sigma_p^{-1} \right) \qquad (2.26)$$

$$\frac{\delta F}{\delta \Sigma_u} = \frac{1}{2} \left[\frac{(u - g(\Phi))^2}{\Sigma_u^2} - \frac{1}{\Sigma_u} \right] = \frac{1}{2} \left(\epsilon_u^2 - \Sigma_u^{-1} \right)$$

where Σ^{-1} denotes the inverse variance of a prediction error. This way of expressing predictive coding models shall be used in the reminder of the thesis, hence the need for a logical progression of 26 equations to lay solid foundations to the rest of the manuscript. As we will see in the implementation section of this introduction, this expression possesses the desirable attribute of requiring only local variables, making it compatible with Hebbian learning rules [126]. Now that we have established a set of robust equations, we can commence our exploration into the crux of the issue: variance terms.

2.1.2.4. On the specific case of variance

A central aspect of predictive coding, which is often overlooked in the literature [254, 72, 293, 337, 213] or reduced to identity matrices, is the variance term in Equation 2.26. As we have seen in the previous pages, this term naturally comes into play in Bayesian inference, and possesses a crucial role in driving the dynamics of a model, driving the posterior distribution of a system closer or further from the likelihood.
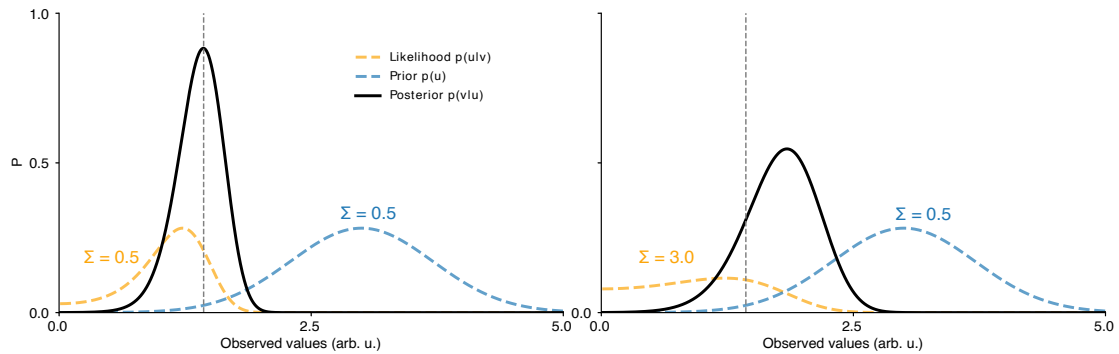
Figure 2.10. – The role of variance in Bayesian integration. (left) As in Figure 2.9, Bayesian integration with $\Sigma_u = \Sigma_p = 0.5$. The most likely value of v is indicated by a gray dashed line. (right) Same but with $\Sigma_u = 6\Sigma_p$, driving the posterior away from the sensory likelihood.

This relates to a number of fascinating emergent properties in neural system which will be discussed in later parts of the introduction, namely of neuromodulation [92], attention [153] and even psychiatric disorders [3, 244]. This also speaks of a hierarchy of prediction and prediction errors, whose multiple integration levels are driven by the relative variance of external inputs and internal predictions. Whilst this is discussed more in details in the implementation part of the introduction, we can already state that low-level prediction errors, like the one encoded in V1, are tightly bound to the variance of the sensory input, and computationally, learning such variance allows to learn about the extrinsic variability of the external world (see chapter 3). This also allows the visual system to factor in the intrinsic variability of its sensors [81, 80].

We shall see in the coming section that the abundance of generic ideas about variance is counterbalanced by the lack of specific neurobiological and neurocomputational literature at the implementational level. As this portion of the introduction is exclusively focused on algorithmic-level concepts, we will not delve deeper into variance at this point. Instead, we will now be focused on computations that can be done on the variance of sensory input, i.e. the likelihood $p(u|v)$.

2.2. The two-fold approach to the problem under study

Having clarified the computational 'why' and algorithmic 'what' of our undertaking, we now turn to the 'how' of its implementation - Marr's final level of analysis, the implementational level, a final stage where our theoretical statements comes to fruition. As previously done, this section aims to provide an overview of the relevant parts of the literature whilst refraining from being a bullet-point detailed list. This is especially the case in the neurobiological section of this introduction, which is concerned with the actual realization of our theories, rather than their experimental aspects, which will be reserved for the introduction of their respective chapters.

2.2.1. The in-silico implementational level: Neurocomputations

The first part of this introduction to the implementational level involves transitioning from the theoretical underpinnings of predictive coding under the free energy principle, towards a practical model that can effectively be employed in vision.

2.2.1.1. Predictive coding for vision

As stated in Equation 2.26, every single variable required by our formulation of predictive coding can be expressed in terms of a graph with local variables. Such network can be represented as in Figure 2.11.
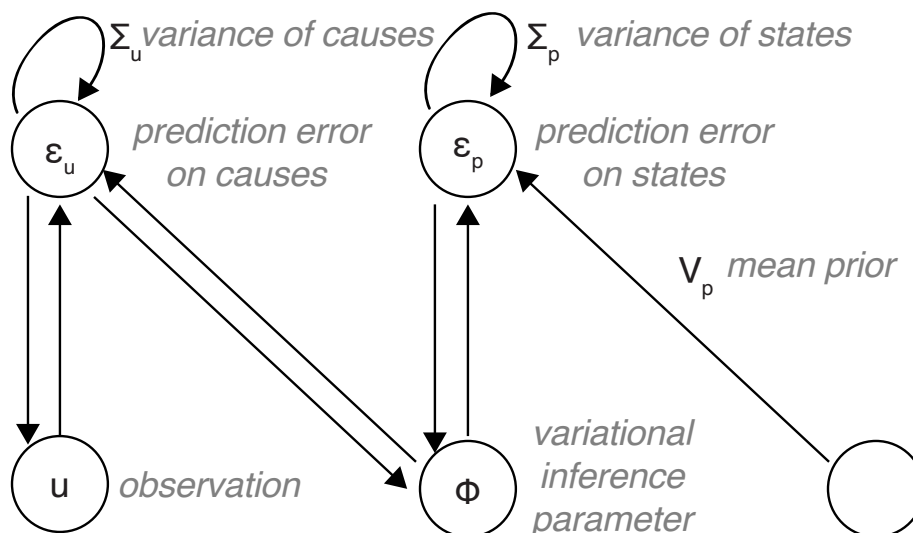
Figure 2.11. – A simple predictive coding graph, with helpful reminder of the variables nomenclature in gray.

This network constitutes what is known as an acyclic computational graph, a generalized method of articulating problems to topology that will prove useful in chapter 5. Empirical studies have also shown that predictive coding can perform on par (or even better) than the main method of training deep neural networks along such a graph [212], a method known as backpropagation [185]. This important property helps us scale such a graph beyond the unidimensional form, as we have been doing so far, to a form where "nodes" of the graphs (neurons) would for example represent multiple different orientation in the visual field, as in Figure 2.12.
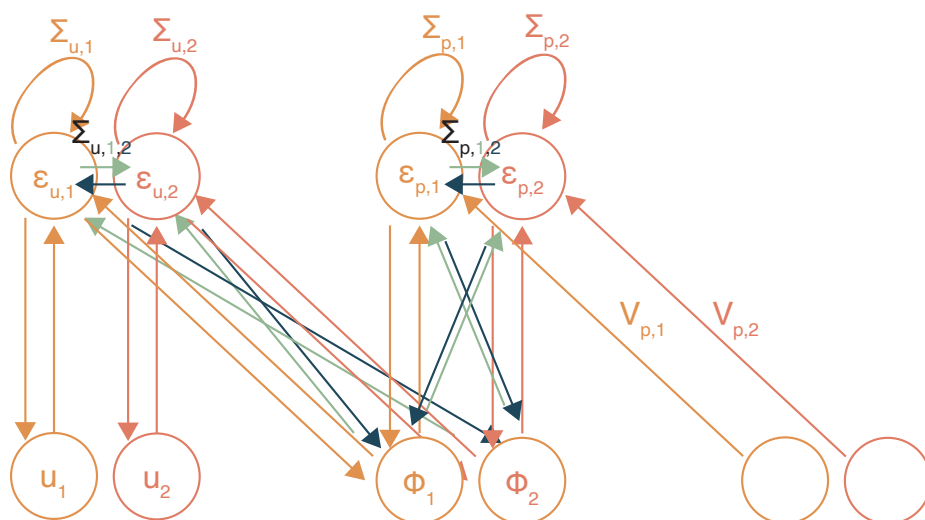
Figure 2.12. – A matrix form (with two elements) of the previous predictive coding graph, with additional inter-matrix elements connections colored in green and dark blue.

Following the derivations of Appendix A, such network then becomes, in matrix notation, where x is a scalar, $\bar{x}$ a vector and $\mathbf{x}$ a matrix:

$$\begin{aligned}
\dot{\bar{\epsilon}}_p &= \bar{\phi} - \bar{v}_p - \Sigma_{\mathbf{p}}\bar{\epsilon}_p \\
\dot{\bar{\epsilon}}_u &= \bar{u} - g(\bar{\phi}) - \Sigma_{\mathbf{u}}\bar{\epsilon}_u
\end{aligned} \tag{2.27}$$

with the dynamics of the matrices being:

$$\begin{aligned}
\frac{\delta F}{\delta \bar{v}_p} &= \bar{\epsilon}_p \\
\frac{\delta F}{\delta \Sigma_{\boldsymbol{p}}} &= \frac{1}{2}(\bar{\epsilon}_p\bar{\epsilon}_p^T - \boldsymbol{\Sigma_p^{-1}}) \\
\frac{\delta F}{\delta \Sigma_{\boldsymbol{u}}} &= \frac{1}{2}(\bar{\epsilon}_u\bar{\epsilon}_u^T - \boldsymbol{\Sigma_u^{-1}})
\end{aligned} \tag{2.28}$$

This approach is extremely useful to us, as it allows us to incorporate multiple

sensory inputs to the network, like multiple oriented edges to V1, which is a critical aspect of the implementational challenge in this thesis. Sadly, it also introduces a major complication, because the inverse variances represented by $\mathbf{\Sigma^{-1}}$ now require matrix inversion. This necessitates a network-wide computation, implying that all nodes must access the entirety of the data instantaneously, which is not biologically plausible. Further, such matrix inversion is computationally demanding [67], which is one reason why inverse variance weighting is mostly absent from existing predictive coding implementations. This now presents us with the first implementational issue related to the computations we have detailed so far: trying to incorporate the algorithms we have designed into a formulation that is biologically plausible. While numerous solutions exist [92, 31, 153], we shall continue one that is naturally suited to matrix forms [31] and involves the addition of an extra inhibitory neuron, as in Figure 2.13.
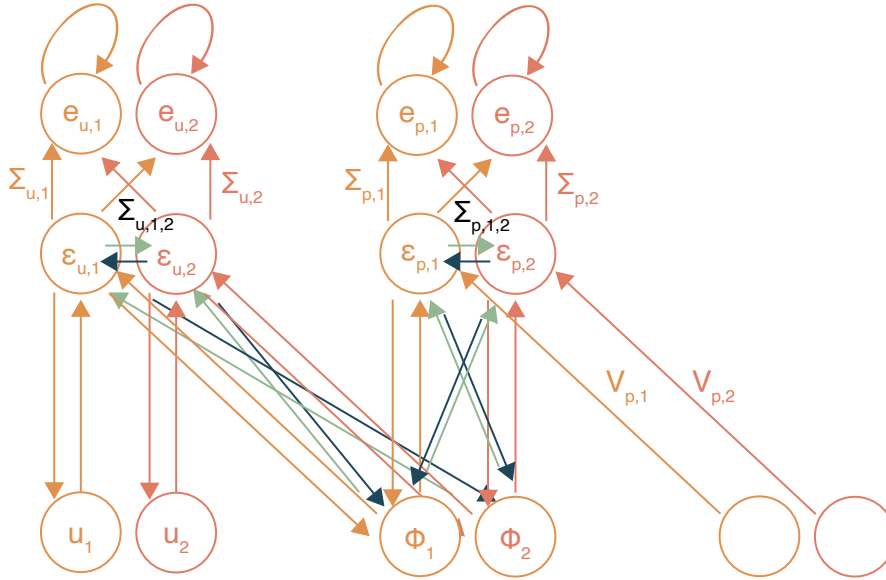
Figure 2.13. – A matrix form predictive coding graph with Hebbian variables.

Based on the notions developed in the final section of Appendix A, the equation of the prediction errors of the network thus becomes:

$$\begin{aligned}
\dot{\bar{\varepsilon}}_p &= \bar{\phi}_p - g(\bar{\phi}_{i+1}) - \bar{e}_p \\
\dot{\bar{e}}_p &= \Sigma_{\mathbf{p}} \bar{\varepsilon}_p - \bar{e}_p
\end{aligned} \qquad (2.29)$$

where e_i represents the additional inhibitory neuron. This implementation answers our needs for being able to handle the uncertain nature of vision through learned Σ matrices, as detailed in the computational part of the introduction, while also being able to transcribe all the equations of the computational part of the introduction into a tangible and useful form.

But where does the "predictive" aspect of this predictive coding stem from ? This type of modelling can be traced back to V1 models developed by Rao and Ballard [254], who transformed what was originally a signal processing algorithm developed for unidimensional signals [196] into arguably one of the most robust theories in neuroscience [98]. By deriving the principle of efficient coding [16], they reasoned that neural networks can all be described as predictive networks, like the one formulated above, and can serve as excellent models for V1. This proved to be groundbreaking, even predicting the existence of non-classical phenomena [254] that previously required dedicated models [288]. The remarkable simplicity and beauty of this approach, carried by the free energy principle, is well captured by a quote from Karl Friston [94]:

> *"Every decade or so, one reads a paper that makes you think "well, that's quite remarkable". [Rao and Ballard] showed that a simple architecture was not only consistent with neuroanatomy and physiology but could also account for a range of subtle response properties [...]*
> *This was a significant achievement in its own right; however, the really remarkable thing —at least for me— was the following: in simulating their little piece of synthetic cortex, neuronal dynamics and connectivity optimized the same energy or cost function."*

This breakthrough sets the stage for the tremendous success of predictive coding in artificial neural networks. Such predictive network perform on par with deep neural networks, can classify images datasets, predict complex natural images sequences, among many other remarkable feats. Unfortunately, the inclusion of variance is often overlooked in the implementation of such networks. This is not only due to the added complexity it brings to an already intricate network, but also because most of these networks employ point-based estimates rather than comprehensive Bayesian-like distribution learning. Such gap is exactly the aim of this thesis.

One exceptional aspect of predictive coding, though not extensively addressed in this thesis, is the fundamental idea that the brain must predict its inputs and possess a generative model of the world to account for its interpretations. What amplifies the significance of this is the fact that these generative models are self-invertible [211]: practically speaking, one can train them, for instance, to classify objects, then simply reverse the flow of information and have them transformed into generative models of images [213]. An example of this reversal can be found at our GitHub repository, detailing the work we conducted at Telluride workshop to use predictive (generative) coding models in order to issue commands to in-vitro neurons.

In light of this series of generative/discriminative model, predictive coding also suggests a hierarchical series of explanations, which aligns remarkably well with the hierarchical nature of visual processes [33, 34]. At the scale of our focus on V1, predictive coding is particularly impactful as there have been numerous attempts to align the computations performed within the cortical "microcolumn" circuits, or microcircuits, with those carried out by predictive coding [20]. This serves as the starting point for our bridge towards the biological aspect of our networks [283], by employing such theories as a lens to delve into the cortex.

2.2.1.2. Microcircuits as a bridge from the theory to the cortex

Having now expressed the problem of probabilistic vision in a form that can be mapped onto a graph, the next step is to transpose that graph into a neural network. In the introductory section related to the neurobiology of vision, we discussed some (controversial [135]) attempts to identify a recurring circuit in the cortex [74, 73]. As a reminder, such "canonical microcircuit" tracks the flow of information through multiple neurons, and was first established through intracellular recording in the cat's V1 [74] to measure pre- and post-synaptic connectivity and functional strength. According to these studies and others [314], a feedforward flow of excitatory activity in the cortical microcolumn arises from thalamic inputs to layer IV, then to layer II/III, and finally back to layer V/VI. There is proof that other excitatory pathways that would close the loop only form a minority of connections, namely the layer II/III to layer IV or layer V to layer II/III [314]. Inhibitory pathways in cortical microcolumn are historically harder to make out, namely because such neurons have much more functional subtypes [116], but recent evidences are showing that they can participate in the regulation of the circuit in a layer-specific fashion [41]. Extrinsically, this circuit needs also to be integrated into an input/output scheme compatible with the idea of hierarchical predictive coding [34]. Inhibitory connections play here an important role, because they allow multiple such microcircuit to compete against one another at the level of a cortical area [64, 50] (see also chapter 4). At the macroscale level, backed by anatomical evidence [199], the final concept of a canonical microcircuit is that layer II/III sends synapses to higher order cortical areas [83], as opposed to layer V/VI that sends synapses to lower levels [198]. Based on these considerations, Bastos et al. [20] proposed a mapping of our computational graph onto a neural network:
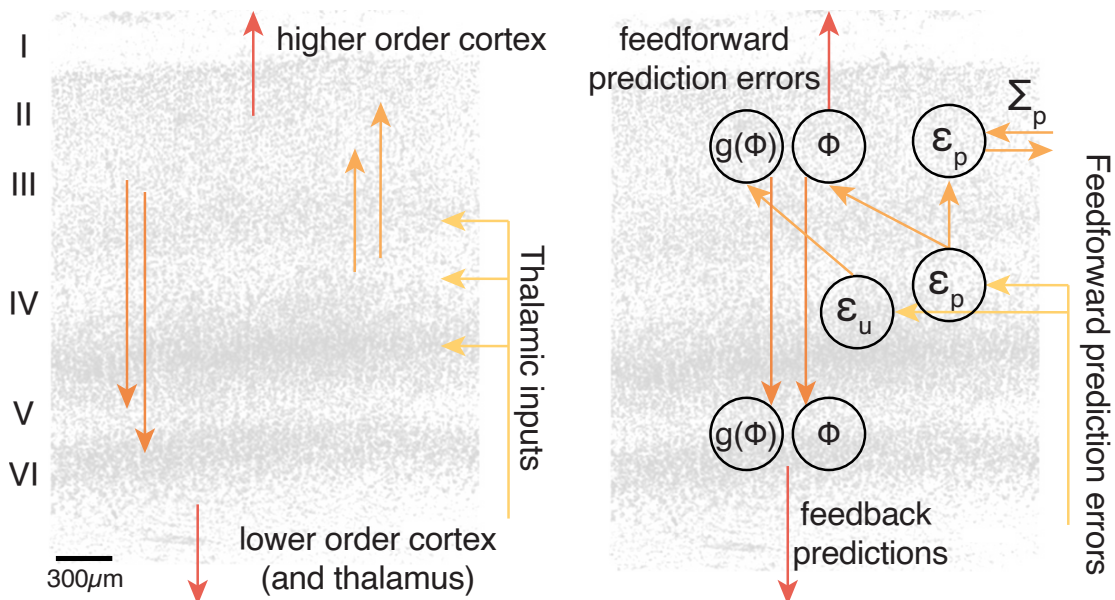
Figure 2.14. – Canonical microcircuit for predictive coding, with (left) Figure 2.4 for comparison and (right) graph adapted from Bastos [20].

It now remains to review experimental proofs to back the existence of such predictive circuitry. In that sense, proofs of a canonical microcircuit are somewhat sparse, but there is growing consensus that a canonical microcircuit performing predictive operations does indeed account for experimental results. In this context, the study in chapter 4 provides further experimental support, by validating the idea of mapping input variance to supragranular layers, largely due to their extensive lateral connectivity [12, 50, 4]. There is also vast account of the fact that specific types of neurons in microcircuits - namely disinhibitory motifs [175, 222] - are performing layer-specific computations [41] that can support learning on-par with predictive coding algorithms [264]. Finally, there is also theoretical evidence that show that combining simple elements of neural origin can actually yield neural networks that compute prediction errors in an unsupervised manner [30, 132].

Further, this model implies that predictions and prediction errors, operating mostly on two separate firing regimes, are reflected in the oscillatory domain (an experimental concept further explored in chapter 7). Based on that observation, there is a mounting evidence that suggest two bands of oscillations for superficial and deep layers of the cortex [22]. Such oscillatory evidences are numerous [227, 270, 77, 76] to show that deep layers oscillate at a low frequency [21] (which is gated through the pulvinar [61], as contributed in chapter 7) to prepare predictions and make way for the fast oscillations of sensory inputs in upper layers.

On a broader scale, numerous empirical evidences of predictive coding mechanisms at work in the brain have been accumulated, even prior to Friston's initial paper, which we will now explore to further support our approach.

2.2.2. The in-vivo implementational level: Neurobiology

The final part of this introduction to the implementational level involves transitioning from the notion that predictive coding can be used as a framework for this thesis, towards showing proof that there exists both predictive and variance-based computations in the brain.

2.2.2.1. Neural evidences of predictive coding

Having gathered the (sparse) supporting biological evidence for a canonical microcircuit, and having mapped this predictive microcircuit to certain functional activity at the circuit level, we now turn our attention towards assessing whether the brain indeed utilizes predictive coding at a global scale. A bias naturally stems here [283], as we will try to explain many disparate phenomena with a single theory - when one perceives every problem as a nail, it becomes convenient to envision a normative theory based on a hammer [203].

Essentially, as we expressed in Equation 2.26, predictive coding proposes that the brain increases computational efficiency by transmitting only prediction errors [92]. As such, if a certain visual input does not generate a prediction error, it should not be transmitted, and thus the neural response for predicted stimuli should be weaker than for unpredicted stimuli. Therefore, the first argument that supports the existence of predictive coding in the brain is the experimental observation that repetition of stimuli elicit a diminution of the evoked neural activity. This is observable not just with the visual responses [304, 305], but also in the auditory [104, 316] and somatosensory cortices. Conversely, unexpected or surprising events should trigger an increase in said neural activity. This assertion is also verified within the visual pathways [237] where the violation of a repetitive pattern induces a significant surge in the firing rate. This effect is also noticeable at the psychophysical level, a neural event known mismatch negativity potential [10, 334], which is an entire sub-field on its own (see chapter 7). Further evidences can be found when violating the distribution of natural images [14, 90] (as described in the first section of the introduction), or under many conditions of predictability violation [170, 207, 221].

It's also worth noting that the large-scale hierarchical organization of the cortex, particularly the visual cortex [83], aligns well with the principles of predictive coding (Figure 2.15). In this representation, the hierarchical organization of the brain emerges from the Bayesian inference process we developed earlier, which relies on a hierarchical conditional model [99]. This corresponds well with the layer-specific responses of the microcircuit and their respective feedforward versus feedback connections [199]. On a larger scale, this constitutes a recurrent circuit that generalizes the formulation we introduced in Equation 2.28: a stable functional model that seeks to minimize prediction errors while continuously updating a prediction-based model of the world.

Still on the anatomical side of the argument, the lateral connectivity intra V1 merits further mention [313, 51, 160], as it provides the support for local competition of orientation-based predictions on the worldly states. This is usually characterized as
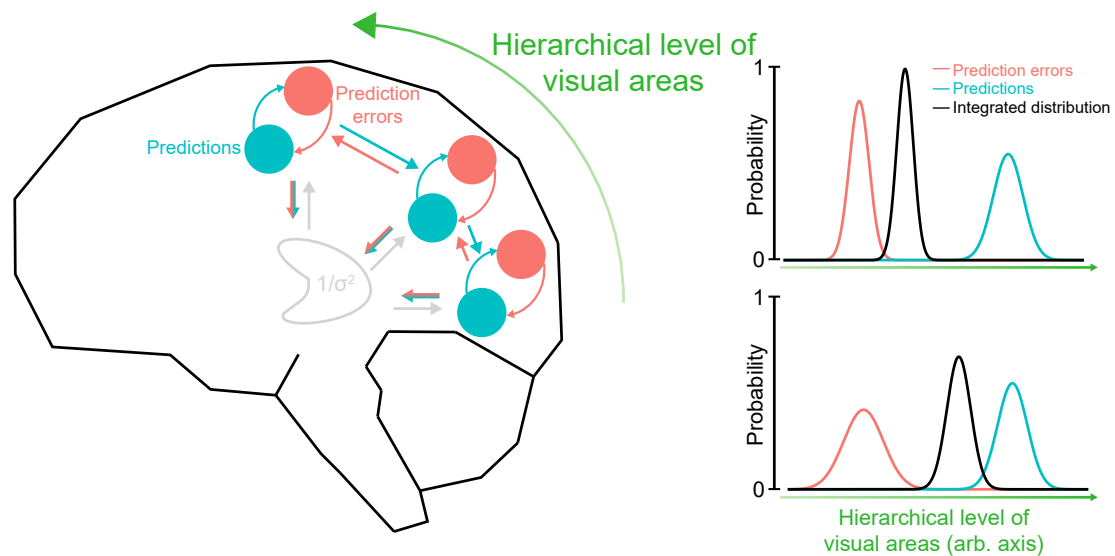
Figure 2.15. – Hierarchical predictive coding in the brain, (left) implemented as a globally and locally recurrent series of prediction errors and prediction integrations. (right) As in Figure 2.10, the effect of the input variance on the predictive model.

an attentional mechanism [93, 6], but we would contribute here that it takes effect far too rapidly [178] for that to be the case. In this thesis, we interpret it as a local mechanism, as formulated by Friston in its original implementation of predictive coding [92]. Another implementation we will discuss in chapter 7 involves a globally parallelized (through the pulvinar) implementation of variance [153], that allows the brain to determine which visual areas best explain the current environment. This is experimentally corroborated by the fact that pulvinar modulations essentially constitute contextual modulations [258, 47, 292] dynamically applying context to the content processed by the cortex [252].

In addition to these findings, predictive coding has been shown to effectively model pathological conditions of the brain, an emerging field referred to as computational psychiatry [3]. While it's too premature to label this as evidence, the wide adoption [117] of this framework is a compelling argument supporting the idea of the brain functioning as a predictive system. In that framework, the idea that shifts in variance can move one closer or further from the prior, as demonstrated in Figure 2.15 (and in the Conclusion of this manuscript), provides a fitting description of disorders characterized by hypo-variant priors (such as autism [323, 324]) or hyper-variant priors (like schizophrenia [134]). This now paves the way for the end of this (lengthy) introduction, with a final (and not lengthy) section dedicated to the representation of variance in the predictive brain.

2.2.2.2. Neural representations of variance

Having mapped predictive models onto some biological substrates, it is now time to explore whether there is empirical biological evidence supporting the representation of variance. Unfortunately, this field of investigation is not an expansive one, as reflected by the length of this section. What sparse evidence exist is however very promising, and in line with our respective contributions in chapters 3 to 8.

The seminal work in this field is a study by Orban et al. [235], which shows that local computations performed between orientation detectors can effectively process orientation variance, as that competition, through inhibition, increases the spike-to-spike variance of the neural activity. Thus, elegantly, the variance of an internal representation can be "signaled" by the variance of the activity supporting it. This aligns well with the psychophysical literature we introduced earlier [128], specifically the notion that competition among orientation detectors accounts for the psychophysical observations of human subjects.

On the biological front, if we refrain from considering the superposition of two gratings as a distribution of orientation, as was historically done [108] (although some approximation hold [102, 190]), the whole idea of encoding orientation variance in V1 was actually pioneered by Goris et al. [115]. They reported that heterogeneously tuned V1 populations help encode the orientation distributions found in natural images, and that this functional diversity could be accounted for by a linear-nonlinear (L-NL) model. While this could explain the diversity of tuning in the data we report in chapter 4, we will find that in terms of modeling, competition among orientation detectors within a predictive context remains the best descriptor [178].
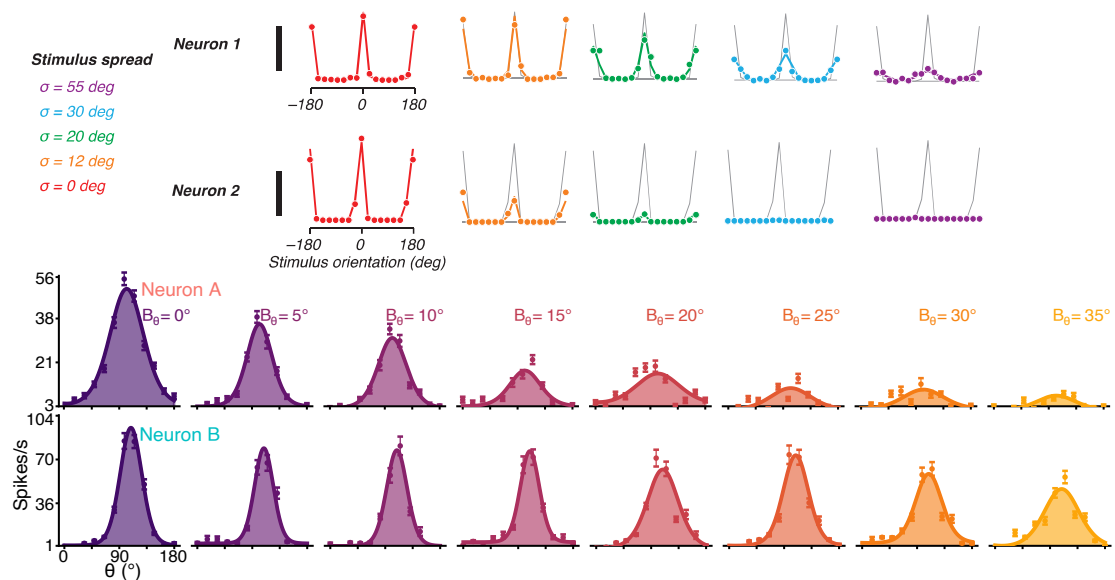
Figure 2.16. – Modulation of tuning curves by orientation variance. (Top) Goris et al. [115]. (Bottom) This thesis [178].

Considering the sparse nature of the relevant literature, referencing our own work

from chapter 4 here feels necessary. Our study reaffirms the findings in the literature on anesthetized macaques [115], as we identified single-neuron variance modulations that underpin the decoding of orientation variance at the population level in V1. This suggests that a shared neural mechanism may exist in both felines and primates, which isn't surprising given the crucial role of variance for the proper encoding of natural images in V1 [230], a point we've emphasized repeatedly throughout this introduction.

Our specific contribution here is connecting these findings to cortical layers, strengthening the idea that supragranular neurons with sharp tuning and slow dynamics [256, 257] facilitate the concurrent encoding of orientation and its variance. This links nicely with the concept of cortical microcircuit introduced earlier, as ten years prior to our article, it was proposed that supragranular activity should encode variance, a hypothesis for which we have provided formal experimental evidence in this thesis.

As far as biological evidences are concerned, there is consensus across studies that heterogeneity and local competition, i.e., intra-V1 activity, are sufficient to explain all observations. In fact, both neurobiological and computational evidence suggest that V1 doesn't need to enlist other cortical areas to process orientation variance, but that such process in other cortical areas might actually be part of synchronized global computations (more in chapter 7).

For instance, the heterogeneous recurrent excitatory and inhibitory synaptic connectivity in V1 [146, 54, 140, 273] sustains resilient orientation tuning [214] that can account for the diversity of single neurons' resilience under different connectivity profiles, as explored in our computational model [178]. This is supported by the temporal scale of local recurrent connectivity, namely the slowly-conducted horizontal waves in an orientation map [51], which fits the view of variance processing as an iterative and accumulative computation implemented by local recurrent interactions between supragranular resilient neurons that are heavily connected through recurrent interactions with neighboring cortical columns [74, 256, 257, 51].

Despite the limited quantity of available evidence, the existing findings notably converge. Whether across species, research teams or functional encoding schemes, the overarching theme remains constant: there is an active encoding of variance in the primary visual cortex. Having established this, we can now proceed to explore the main section, beginning with the exploration of the structure of variance in natural images, which will then directly link to the findings in chapter 4.

3. Variance in Vision Models: a Convolutional Sparse Coding Approach

"These stats are staggering
Had his Ph.D in indiscreet street haggling."
MF Doom, Gazzillion Ear, 2009

Summary

3.1. Introduction: Orientation, Statistics, and Orientation Statistics

As introduced in the previous chapter, a central role of the brain is to build a model of its environment in order to enact complex behaviors. One key component of these models is their reliance on internal representations of the environment. While this is true in predictive coding (where these representations are predictions), it also forms the dominant narrative in neuroscience [163]. This understanding of the brain stems from the discovery of the receptive field, by Sherrigton's seminal studies [281], which correlated a neural discharge pattern with an element of the environment (here, stroking a certain skin spot). Iteration upon iteration of related research have built modern neuroscience upon a similar narrative, starting with the "fly detectors" neurons of Barlow [17], the orientation selective neurons of Hubel and Wiesel [137], the place cells of O'Keefe and Dostrovsky [229], the grid cells of Hafting [122], the face cells of Perret [243], and so on [201].

In vision, there is thus no denying that orientation forms the basis of these models of the world internalized in the brain. One legitimate question that should be asked

before diving into 4 chapters of studies related to orientation selectivity could be: "why orientations in the first place" ? As history would have it, Hubel and Wiesel discovery was somewhat serendipitous [201], and orientation selectivity was discovered somewhat accidentally when the two scientists noticed strong neural responses as they were inserting a glass slide (used to project the spots of light) into the projector of their cat experiments. It thus become evident that neurons in V1 were in fact responding not to spots of light, as in the retina and LGN, but to the edge of the glass slide, indicating a selectivity for oriented edges.

Had computational neuroscience been 30 years more advanced at the time of these experiments, Hubel and Wiesel might have had the chance of knowing what to look for in the first place, rather than stumble upon it semi-accidentally. Indeed, we have said that a key property of the brain is efficient coding [18], which saves costly neurobiological message passing by encoding solely relevant information. Under predictive coding, for example, this means solely transmitting prediction errors. This generic principle imposes a constraint of sparseness on the message to be sent by neurons, meaning using as little energy as possible while maintaining a highly accurate internal model. Thus, one can envision early sensory cortices as models of the world build through the transformation of dense redundant inputs into sparse efficient representations. By creating a computer model that performs high-quality reconstruction with as little activity as possible, one can thus see what types of features are ideal to deconstruct any given type of sensory input. This trade-off has been explored by Olshausen and Field [230], who showed that a model of natural images with a constraint of sparsity yields receptive fields that are extremely similar to those found in V1 (Figure 3.1). Thus, a low-level invariant representation of the (static) visual world can be created using edges detector.
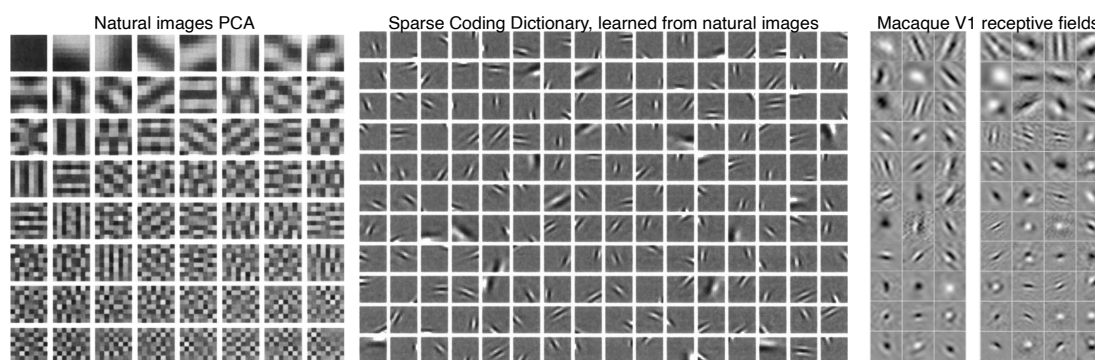
Figure 3.1. – Natural images and sparse dictionaries. (left) Natural images components, extracted using Principal Component Analysis (PCA). (middle) Dictionaries yielded by learning a sparse code on natural images, from [230]. (right) Comparison with macaque V1 receptive fields, from [257].

The statistical distribution of these edges in any given image follows a characteristic pattern [245], and given optimal modelling, these statistics are the main constraint upon which further sensory processing relies [232]. Most of these oriented edges are

represented at the cardinals points, that is, horizontally and vertically [60]. This is echoed at the neuronal level by a cardinal bias in visual perception [124]. As we will show in this chapter's article, around either of these main orientations, the distribution of oriented elements can be characterized by its first- and second-order moments: a median orientation, and its corresponding (inverse) variance. A proper model of a natural images thus depends on a proper model of both these moments (Figure 3.2), which is reflected in the response properties of primary visual cortex neurons that have diverse orientation and variance [138].
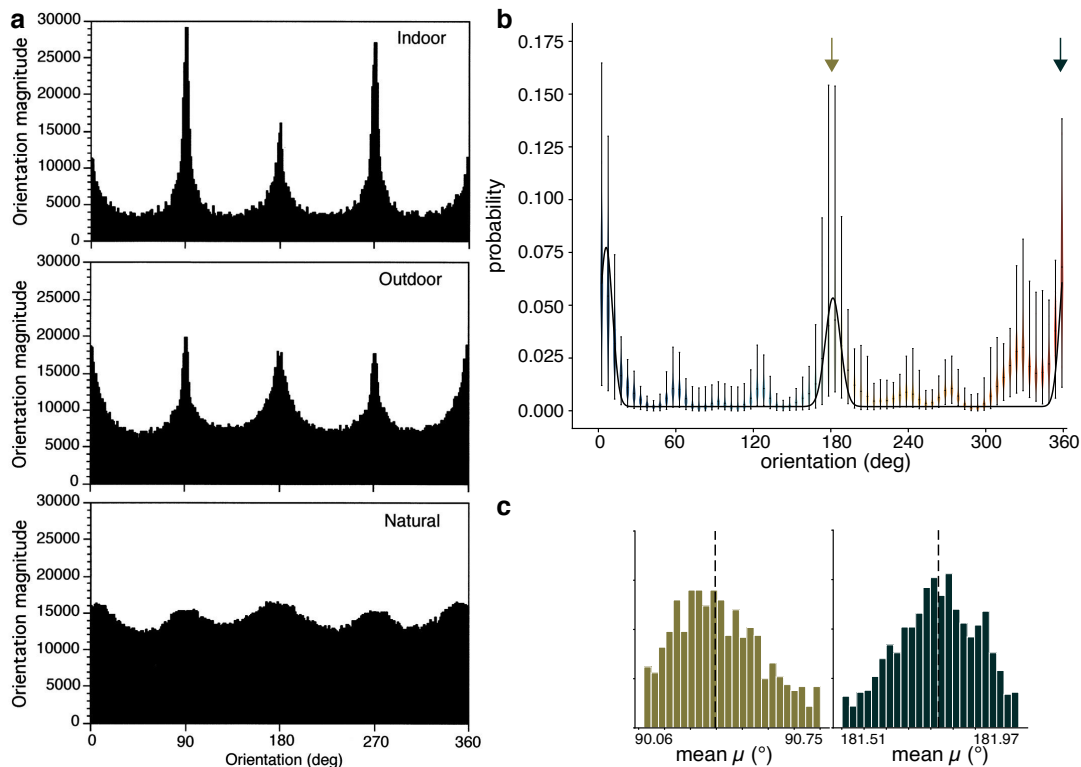
Figure 3.2. – Distribution of orientations in natural images. (a) Distribution of oriented contours, extracted by Sobel filters (see this chapter's article), as done in [60]. (b) Distribution of oriented contours, extracted by Sparse Coding (as done in this chapter's article), with mean von Mises distribution in black. (c) Distribution of means of two peaks of the distributions, for each image.

If orientations in natural images follow such a prototypical distribution, then what is the optimal neural code to represent them ? This question is one of uncertainties. Uncertainty on how the image's orientation deviates from the typical distribution is a problem of uncertainty bound to the input, called aleatoric uncertainty. Uncertainty on how to best model these images is a problem of uncertainty bound to the model, referred to as epistemic uncertainty. Aleatoric uncertainty is linked to the variance of the distribution of orientation: the higher the variance, the more spread the input is (in orientation space), and thus the less certain the information is. As stated in

the introduction, linking this aleatoric uncertainty/variance to the uncertainty of the model is crucial for Bayesian processing, which provides explicit rules to do so (see Equation 2.3).

This chapter serves to characterize natural images as Gaussian (or here, von-Mises) distributions of oriented features, which allows us to keep working within the mathematical framework described in the introduction. This further serves as a justification of Gaussian distribution of orientation as stimuli for animal recordings in chapter 4. Second, it serves to understand how the variance of natural images can be processed by a model with no explicit learning rules for variance. This, again, is useful to describe variance interactions as an emergent property throughout this thesis. Third, as we'll delve into in this chapter's conclusion, this perspective enables us to examine the trade-off between two encoding strategies. One approach involves dense sampling using numerous neurons, providing high quality reconstruction but at a higher energy cost. The alternative employs sparse sampling, where fewer neurons co-encode features and their variance. Under the right conditions, this latter strategy can be equally performant, but much more energy-efficient.

3.2. Methods: Sparse Coding

The model used by Olshausen and Field [230] to show the emergence of orientation selectivity in sparsely-constrained algorithm is referred to as Sparse Coding (SC), and is a widely used model for learning the inverse representation of an input signal. Given the assumption that a signal can be represented as a linear mixture of basis functions (in neurobiological terms, receptive fields), the optimization problem solved by sparse coding is one that tries to minimize the number of basis functions that are used to represent the input signal (which would be spikes), yielding a compact and efficient representation of the original signal. Here, we framed sparse coding as the problem of reconstructing an image s from sparse representations x while minimizing a L_1 norm of the representation. This problem can be approached with a Basis Pursuit DeNoising (BPDN) algorithm:

$$\underset{x}{\operatorname{argmin}} \frac{1}{2} ||s - Dx||_2^2 + \lambda ||x||_1 \tag{3.1}$$

where D is a dictionary (i.e. a set of basis functions used to represent s) and λ a regularization parameter that controls the trade-off between fidelity and sparsity. The present article uses a variation of sparse coding, Convolutional Sparse Coding (CSC), which, as the name implies, relies on the convolution operator:

$$\underset{\{x_k\}}{\operatorname{argmin}} \frac{1}{2} ||s - \sum_{k=1}^{K} \mathrm{d}_k * x_k||_2^2 + \lambda \sum_{k=1}^{K} ||x_k||_1 \tag{3.2}$$

where x_k is an N^2 dimensional coefficient map (given an N^2 sized image), d_k is one kernel (among K channels) and $*$ is the convolution operator (Figure 3.3). One addi-

Figure 3.3. – (a) Convolutional Sparse Coding (bottom) relies on convolution operations to achieve global sparseness and remove local redundancies (top), mimicking the operation of the visual system. (b) and (c), visualized as vector-vector or vector-matrix products.

tional advantage of convolutional sparse coding over other reconstruction techniques is its ability to learn interpretable features from the data, which can be easily visualized and understood by human experimenters [35].

3.3. Methods: Deep Learning

The final section of this article uses a Deep Neural Network, as a model to understand how the sparse code can be of use in further stage of visual hierarchical processing. While they are not used elsewhere in this thesis, a brief introduction can be useful. All modern artificial neural networks rely on the gradient descent algorithm, as introduced by Rumelhart [262]. The essence of this optimizer is to iteratively adjust the parameters of a neural network to minimize a loss function, defined as the difference between its predictions and the actual data.

Given this loss function $L(\theta)$, where θ represents the parameters of the network, the gradient descent update rule is expressed as:

$$\theta_{t+1} = \theta_t - \alpha \nabla L(\theta_t) \tag{3.3}$$

where θ_{t+1} is the updated parameter at iteration $t+1$, α is the learning rate, and $\nabla L(\theta_t)$ is the gradient of the loss with respect to the parameters at iteration t.

Here, we use a Convolutional Neural Network, as introduced by LeCun et al. [186] to demonstrate the capability of learning hierarchical features is a central model in Deep Learning for processing grid-like topology data, such as images. This model is based on the idea that an input signal can be represented through a hierarchical set of layers, where each layer transforms the input data with the aim of gradually abstracting the features of the data to enable effective classification or regression at the output layer.

Given an input image I, a Convolutional Neural Network seeks to learn a hierarchy of convolutional features F by applying a series of convolutional and pooling operations, typically defined as:

$$F_l = \text{ReLU}(W_l * F_{l-1} + b_l) \tag{3.4}$$

where F_l is the feature map at layer l, W_l is the convolutional kernel, b_l is the bias term, and $*$ is the convolution operator.

Deeper architectures, that is, with more layers, also contain a MaxPooling operation which groups representations into an intermediate, dense form:

$$F_{\text{deep}} = \text{MaxPooling}(\text{ReLU}(W_{\text{deep}} * F_{\text{prev}} + b_{\text{deep}})) \tag{3.5}$$

where F_{deep} represents the deeply learned features, and F_{prev} is the feature map from the previous layer. These Deep Convolutional Neural Networks, with their deep architectures, have an advantage over other models due to their capability to learn more abstract and generalized representations of input data, which are critical for solving complex problems in computer vision [173]. In this study, we leverage this capability to understand how the representation power of deep layers influences the

performance of the model. Each DCNNs has its particular architectural tweaks, for which we would refer the reader to the article in the next section.

3.4. Article: "Sparse Representation of Natural Images with Heterogeneous Orientation Kernels"

The following article is the initial contribution of this thesis, framing natural images as Gaussian-like distribution of oriented elements. This alleviates major non-linear integration difficulties that would otherwise be present in the predictive coding framework, and serves as a justification of the use of Motion Clouds [189] in the next chapter. On its own, the article uses dictionaries of receptive fields that emphasize two coding strategies to reconstruct natural images: focusing either on median features (orientations) or their variance (bandwidths). We show that focusing on the former improves reconstruction, while the latter improves sparseness. Fine-tuning through learning on a dataset of natural images alleviates this compromise, allowing optimal encoding of natural images through a sparse co-encoding strategy, as will be also uncovered in V1 in chapter 4.

Full citation is as follows: Hugo J Ladret, Christian Casanova, and Laurent Udo Perrinet. "Kernel Heterogeneity Improves Sparseness of Natural Images Representations". In: *arXiv preprint arXiv:2312.14685* (2023)

1   Introduction
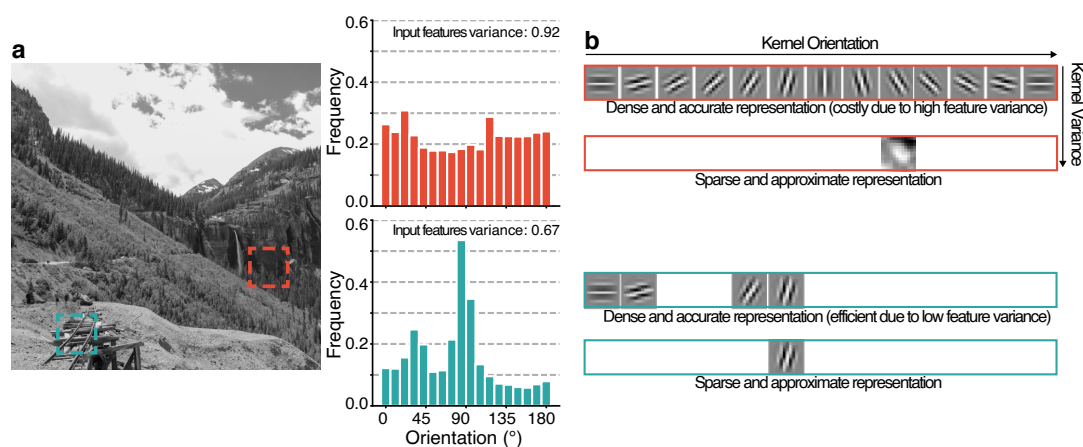
Figure 1:   Efficient coding of sensory inputs. **(a)** Orientation distributions with high (red) and low (blue) variance, in two 256^2 pixel patches from a sample natural image. **(b)** Representation of these distributions and their efficiency depends on the structure of the input. The high-variance patch can be accurately represented with multiple oriented kernels, or approximated using one single kernel with high representational variance. Similarly, the low-variance patch can be encoded as a two-peaked orientation for an accurate representation, or using one kernel of low representation variance for a higher sparseness.

Neuromorphic neural networks are fundamentally designed to process inputs based on their statistical characteristics. This is particularly evident in vision-related tasks related to natural images, which exhibit a set of common statistical properties at multiple levels of complexity [1]. These statistical characteristics guide sensory processing, and are implicitly learned through efficient coding models [2, 3]. For example, natural images typically show a local redundancy in luminance patterns that biological neural network remove at early processing stages, enhancing computational efficiency [4]. In general, these images can be conceptualized as distributions of features (Figure 1), which are, at a low descriptive level, oriented edges that form the foundation of hierarchical representations of natural images [5]. The first moment of these distributions informs on the mean orientation in a given image patch, while the second central moment represents the heterogeneity of these features.

Modeling of such heterogeneity is crucial for sensory processing, both through input and representation bound variances [6]. Input variance, also referred to as *aleatoric* variance, stems from the intrinsic stochasticity in the processes that generate natural sensory inputs, such as sounds [7], textures [8] or images [9]. As its sources escape modeller control, it is challenging to predict, especially in computer vision models [10] or neuromorphic hardware [11], and mandates a robust approach to accurately represent and process naturalistic inputs.

Evidences from neurobiological networks support the notion that neural systems account for this variance in decision-making processes [12], following Bayesian-derived

rules [13]. In practice, this is supported through the variability of neuronal sparse activations [14], which depends directly on the variance of the input [15, 16]. This relationship ties input variance to representational variance : in feature space, the basis function of a neuron is intrinsically linked to its capacity to encode particular levels of aleatoric variance [17]. Neurons with broad kernels will more effectively encode broadly represented elements in orientation space, such as textures (see Figure 1). This neurobiological evidence can notably serve to "explain away" irrelevant input to neural networks, thereby optimizing neuromorphic designs at the hardware level.

Indeed, neuromorphic machine learning models which emulate the visual system, such as sparse coding, exhibit a dictionary of kernels which possess a wide range of tuning heterogeneity [3]. This heterogeneity is particularly notable in their convolutional forms, where feature activations, being both position- and scale-invariant, effectively mirror the aleatoric structure of natural images. This process is akin to maximum likelihood estimation, wherein modeling visual inputs involves capturing the variance of visual features through parametrized surrogate distributions. Thus, sparse coding, with its minimalistic yet effective neuromorphic approximation of the early visual system, provides a valuable theoretical framework for understanding how input variance is tied to representational variance.

Here, we aim to provide an empirical account of this relationship, namely by showcasing the advantages of incorporating kernels with heterogeneous feature representations in sparse coding models of natural images. We use a convolutional sparse coding model, trained to reconstruct a novel dataset of high-definition natural images, and manipulate the heterogeneity of its kernels to study its reconstruction performances. We show that optimal learning relies on balancing the heterogeneity of features, which reflects the aleatoric variance in natural images. In a general context, we provide a full PyTorch implementation of our convolutional sparse coding algorithms, and use these codes as inputs of a deep convolutional network, boosting resilience to adversarial input degradation. This underscores our finding that inherent heterogeneity of kernels in machine learning, akin to that of receptive fields in biology, enhances computational efficiency by effectively mirroring the statistical properties of inputs.

2 Methods

2.1 Convolutional Sparse Coding

Sparse coding (SC) is an unsupervised method for learning the inverse representation of an input signal [18]. Given the assumption that a signal can be represented as a linear mixture of kernels (or basis functions), SC aims to minimize the activation of kernels used to represent the input signal, yielding an efficient representation [19] that can be inverted for reconstruction. Here, SC was used to reconstruct an image s from sparse representations x, while minimizing the ℓ_1-norm of the representation:

$$\underset{x}{\mathrm{argmin}} \frac{1}{2}||s - Dx||_2^2 + \lambda ||x||_1 \tag{1}$$

where D is the set of kernels used to represent s (called a dictionary) and λ a regularization parameter that controls the trade-off between fidelity and sparsity. Conveniently,

this problem can be efficiently approached with a Basis Pursuit DeNoising (BPDN) algorithm [20]. As there is *a priori* no topology among elements of the dictionary, SC does not preserve the spatial structure of the input signal, which can be problematic in the context of the representation of natural images. Moreover, the overall decomposition is applied globally and handles poorly the overlap between redundant statistical properties of patches in the image [1], yielding a suboptimal representation of the input signal [21].

These problems are leveraged by Convolutional sparse coding (CSC), an extension of the SC method to a convolutional representation, which is closer to a rough neurally-inspired design [22] as used in deep convolutional network (CNNs) [23]. These CNNs use localized kernels similar to the receptive fields of biological neurons in the primary visual cortical areas. A convolutional architecture uses convolutional kernels (dictionary elements) that are spatially localized and replicated on the full input space (or possibly with a stride which subsamples that space). The number of kernels in the dictionary defines the number of features, or *channels*. In CSC, the total number of kernels with respect to standard SC is multiplied by the number of positions. As a result, a convolution allows to explicitly represent the spatial structure of the signal to be reconstructed. This further reduces the number of kernels required to achieve an efficient representation of an image, while providing shift-invariant representations. CSC extends equation (1) to:

$$\underset{\{x_k\}}{\operatorname{argmin}} \frac{1}{2}||s - \sum_{k=1}^{K} \mathrm{d}_k * x_k||_2^2 + \lambda \sum_{k=1}^{K} ||x_k||_1 \tag{2}$$

where x_k is a N^2 dimensional coefficient map (given a N^2 sized image), d_k is one kernel (among K channels) and $*$ is the convolution operator. As the convolution is a linear operator, CSC problems can be solved with convolutional BPDN algorithms [24]. Here, we used the Python SPORCO package [25] to implement CSC methods, using an Alternating Direction Method of Multipliers (ADMM) algorithm [26] which splits Convolutional Sparse Coding problems into two alternating sub-problems, as described in Appendix A. Additionally, CSC proves advantageous over other reconstruction techniques in its ability to learn interpretable and visualizable kernels from input data.

2.2 Dictionaries

Optimal dictionaries to reconstruct natural images are known to be localized, oriented elements [27, 2]. Here, we utilized log-Gabor filters, which have been shown to accurately model the receptive fields of neurons in the visual cortex. These filters have several advantages compared to Gabor filters, notably that they do not have a DC component and that they optimally capture the log-frequency structure of natural images to ensure its optimal reconstruction [28]. The log-Gabor filter [29] is defined in the frequency domain by polar coordinates (f, θ) as:

$$G(f, \theta) = \exp\left(-\frac{1}{2} \cdot \frac{\log(f/f_0)^2}{\log(1 + \sigma_f/f_0)^2}\right) \cdot \exp\left(\frac{\cos(2 \cdot (\theta - \theta_0))}{4 \cdot \sigma_\theta^2}\right) \tag{3}$$

where f_0 is the center frequency, σ_f the bandwidth parameter for the frequency, θ_0 the center orientation and σ_θ the standard deviation for the orientation. This provides with a

parametrization of the dictionary, which is useful to compare the efficiency of different sparse coding models [30]. We kept $f_0 = \sigma_f = 0.4$ cpd, varying only the orientation-related parameters to build the dictionaries. The angular bandwidth B_θ of the log-Gabor filter, expressed in degrees, was defined as $B_\theta = \sigma_\theta \sqrt{2 \log 2}$ [31].

To titrate the impact of including heterogeneity in the dictionary, we created two log-Gabor dictionaries with the same number of channels, one with homogeneous (a single σ_θ) the other with heterogeneous (multiple σ_θ) variance of representations. We compared these dictionaries before and after fine-tuning on the dataset, using a dictionary learned from scratch over the dataset as a fifth reference. Such learning was done by performing convolutional sparse coding in a multi-image setting:

$$\underset{\{x_{k,j}\}}{\text{argmin}} \frac{1}{2} \sum_{j=1}^{J} || \sum_{k=1}^{K} d_k * x_{k,j} - s_j ||_2^2 + \lambda \sum_{k}^{K} \sum_{j}^{J} ||x_{k,j}||_1 \text{ s.t. } \forall k, ||d_k||_2 = 1 \quad (4)$$

where s_j is the j-th image in the dataset and $x_{k,j}$ is the coefficient map for the k-th filter and the j-th image. This was alternated with an optimization step of the dictionary:

$$\min_{D} \sum_{i=1}^{N} \frac{1}{2} \|x_i - D * z_i\|_2^2 \quad (5)$$

subject to the constraint $|d_k|_2 \leq 1$ for $k = 1, \ldots, K$.

Performance of these dictionaries was measured with two metrics. The peak signal-to-noise ratio (PSNR), a common metric to evaluate reconstruction quality of grayscale images, is defined as:

$$\text{PSNR}(I_1, I_2) = 20 \cdot \log_{10}(\max(I_1)) - 10 \cdot \log_{10} \left(\frac{1}{m \cdot n} \sum_{i=1}^{m} \sum_{j=1}^{n} (I_1 - I_2)^2 \right) \quad (6)$$

where $\max(I_1)$ is the maximum pixel intensity of the source image. The right hand-side term of the PSNR is the $\log_{10}$ of the mean squared error, where I_1 and I_2 represent the pixel intensity in the source and reconstructed images, respectively. Given that the natural images used here are encoded on 8 bits, common values of PSNR range between 20 (worse) to 50 (best) dB. We also measured the sparseness of the algorithm, which was defined as the fraction of basis coefficients used in a reconstruction which are equal to zero. This value is between 0 (no nonzero coefficient) and 1 (all coefficients are zero). Parametrization of the algorithm was chosen to balance sparseness and PSNR (Appendix A), i.e. $\lambda = 0.05$, with 750 iterations of the learning phase, a residual ratio of 1.05 with relaxation at 1.8, and dictionaries with $K = 144$ total elements of 12^2 pixels each.

2.3 Histogram of oriented gradients

The distributions of oriented features in Figure 1 were computed using a histogram of gradient orientations. Using the 'scikit-image' library [32], given an input image I of dimension $M \times N$, two gradients were computed at each pixel using Sobel filters

$G_h(x, y)$ and $G_v(x, y)$, respectively, for vertical and horizontal gradients. The maps of the magnitude G_m and direction θ were then given as:

$$G_m(x, y) = \sqrt{G_h(x, y)^2 + G_v(x, y)^2}$$
$$\theta(x, y) = \arctan 2(G_v(x, y), G_h(x, y))$$

$$(7)$$

The range of possible gradient directions over $[0, \pi]$ was divided into 18 bins. The orientation histogram H for each bin b was computed as:

$$H(b) = \sum_{(x,y)} I_b(\theta(x, y))$$

$$(8)$$

where I_b is an indicator function, ranging from 1 if $\theta(x, y)$ falls within the range of the bin b and 0 otherwise. In that context, one can quantify the orientation content in natural images, then estimate the distribution of oriented features within the input: aleatoric variance can then be approximated as the inverse of the squared variance of this distribution in orientation space and is computed as $\text{Var}_{\text{circ}} = 1 - \sqrt{\bar{X}^2 + \bar{Y}^2}$, where $\bar{X}$ and $\bar{Y}$ are the average cosine and sine values respectively, yielding a scalar value between 0 (lowest orientation variance) and 1 (highest).

2.4 Dataset

Images for the CSC sections were captured using either a Canon EOS 650D or Canon EOS 6D camera, fitted with 28mm lenses. A total of 1145 images was collected at a resolution of at least 5184×3456 pixels. For CSC, we extracted and used the central 256×256 pixel segment of each image. These images represent a variety of dynamic scenarios, and were carefully shot to ensure that the subjects of interest were in focus and entirely within the frame. We have made this dataset publicly available on Figshare [33].

2.5 Image classification using deep learning

To evaluate the role of sparse codes obtained, we decided to go further than only measuring representation performance by applying these codes on a common machine learning task: image classification. To perform such classification in a neuromorphic-inspired setting, we utilized a modified version of the CIFAR-10 dataset. This dataset, which is commonly used for image classification, originally contains $60,000$ color images of 32×32 pixel resolution across 10 balanced classes. We processed these images by first upscaling them to 128×128 resolution via bilinear interpolation. Subsequently, they were converted to grayscale and sparse-coded, as described above.

The dataset was divided into a training set containing $50,000$ sparse codes and a test set comprising $10,000$ sparse codes. The network was trained from scratch through a standard PyTorch implementation, with backpropagation of the gradient using the Adam optimizer [34]. The training objective was to minimize the categorical cross-entropy loss, defined as:

$$J(\theta) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{C} y_{ij} \log(\hat{y}_{ij})$$

$$(9)$$

where N is the number of samples, C is the number of classes, y_{ij} is the true label, and $\hat{y}_{ij}$ is the predicted label. The Adam update rule for each parameter θ is based on moment estimates given by:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{10}$$

where η is the learning rate, $\hat{m}_t$ and $\hat{v}_t$ are estimates of the mean and variance of the gradients, and ϵ is a small constant to prevent division by zero.

The sparse codes representing these images were then used as inputs for an adapted ResNet-18 architecture [35] which is a classically used CNN architecture. This deep residual neural network, typically composed of 18 layers and used for various vision tasks, was adapted to process the 144 dimensions of the sparse-coded inputs instead of the standard 3-channel (RGB) format. This dimensionality corresponds to the number of channels in our sparse coding dictionary. No other modifications were implemented in the network architecture design.

Hyperparameters were tuned via grid search to maximize accuracy on heterogeneous variance codes, with the resulting values: $\eta = 2e - 4$, $\hat{m}_t = 0.9$, $\hat{v}_t = 0.99$, $\epsilon = 1e - 08$. When training the network, CSC methods using ADMM algorithms were ported from SPORCO to a custom PyTorch implementation (available at `https://github.com/hugoladret/epistemic_CSC`) to speed up computations.

3 Results

3.1 Heterogeneous kernels improve the sparseness of natural images representations

We explored how variance in sensory inputs and neuromorphic representations controls the encoding strategies of natural images. We compared five distinct convolutional sparse coding dictionaries of similar sizes. Two dictionaries using Log-Gabor filters were constructed : one with a homogeneous level of orientation variance ($B_\theta = 12.0°$) and 72 orientations θ_0 ranging from 0° to 180° (Figure 2a, green) compared to another one with heterogeneous orientation variance, spanning 12 orientation values θ_0 and six B_θ ranging from 3° to 30° (Figure 2b, blue). We then benchmarked these constructed dictionaries against their learned counterparts, which were fine-tuned on the dataset (Figure 2a, orange; b, purple). A final comparison was made against a randomly initialized dictionary learned *de novo* on the same dataset (Figure 2c, black). Performance evaluation across the $1,445$ high-definition natural images revealed that dictionaries initialized with Log-Gabor filters consistently displayed highly variant performance from image to image (Figure 2d). Prior to learning, the dictionary integrating heterogeneous orientation variance outperformed its homogeneous counterpart in sparsity (Mann-Whitney U-test, $U = 1310760.0$, $p < 0.001$), but had significantly lower PSNR ($U = 262261.0$, $p < 0.001$). Post-learning, all dictionaries had similar performances in terms of both sparsity ($U = 634605$, $p = 0.18$ for homogeneous vs random initialized dictionaries ; $U = 634605.0$, $p = 0.97$ for heterogeneous vs random initialized dictionaries) and PSNR ($U = 694175$, $p = 0.46$; $U = 653943.0$, $p = 0.99$). This
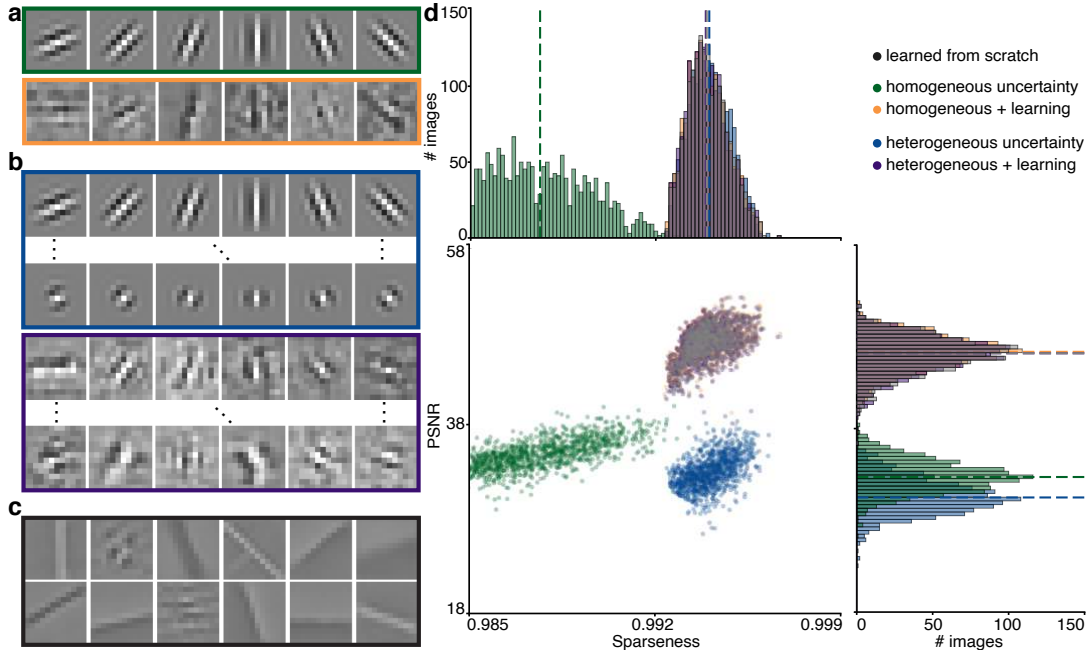
Figure 2: Kernel heterogeneity and reconstruction trade-off. **(a)** Elements from dictionaries with homogeneous kernel variance before (green) and after dictionary learning (orange). **(b)** Same, with heterogeneous kernel variance before (blue) and after learning (purple). **(c)** Elements from a dictionary learned from random initialization on the dataset. **(d)** Distribution of the sparseness (top) and Peak Signal-to-Noise Ratio (PSNR, right) of the five dictionaries. Median values are shown as dashed lines. All three post-learning dictionaries have overlapping (but not identical) distributions.

suggests that emphasis on heterogeneous variance modelling improves the sparsity, at the cost of reconstruction performance.

After learning from the dataset, whether from random initialization or from a pre-constructed log-Gabor dictionary, all dictionaries converge to qualitatively quite different filters, yet with a similar, superiorly sparse and performant form of encoding. The learning method indeed enhanced all Log-Gabor dictionaries, resulting in increased PSNR ($U = 0.0$, $p < 0.001$; $U = 181535.0$, $p < 0.001$, homogeneous and heterogeneous variance dictionaries, compared to their pre-learning version) and sparseness ($U = 23595.0$, $p < 0.001$; $U = 248667.0$, $p < 0.001$). Given the converging reconstruction and sparseness for all these dictionaries, we now focus on the heterogeneous variance dictionary, both pre- and post-learning, as well as the pre-learned homogeneous variance dictionary. Additional performance details for the homogeneous dictionary are provided in Appendix B.

What are then the kernel features changed through the learning process? While fine-tuned dictionaries do incur a significantly higher computational cost during the learning phase, they deliver substantial improvements in both PSNR and sparsity, compared to merely introducing heterogeneous variance into a pre-existing dictionary. These enhancements can be attributed to modifications in the dictionary coefficients following the learning phase, affecting both the feature orientations (θ_0) and their associated lev-
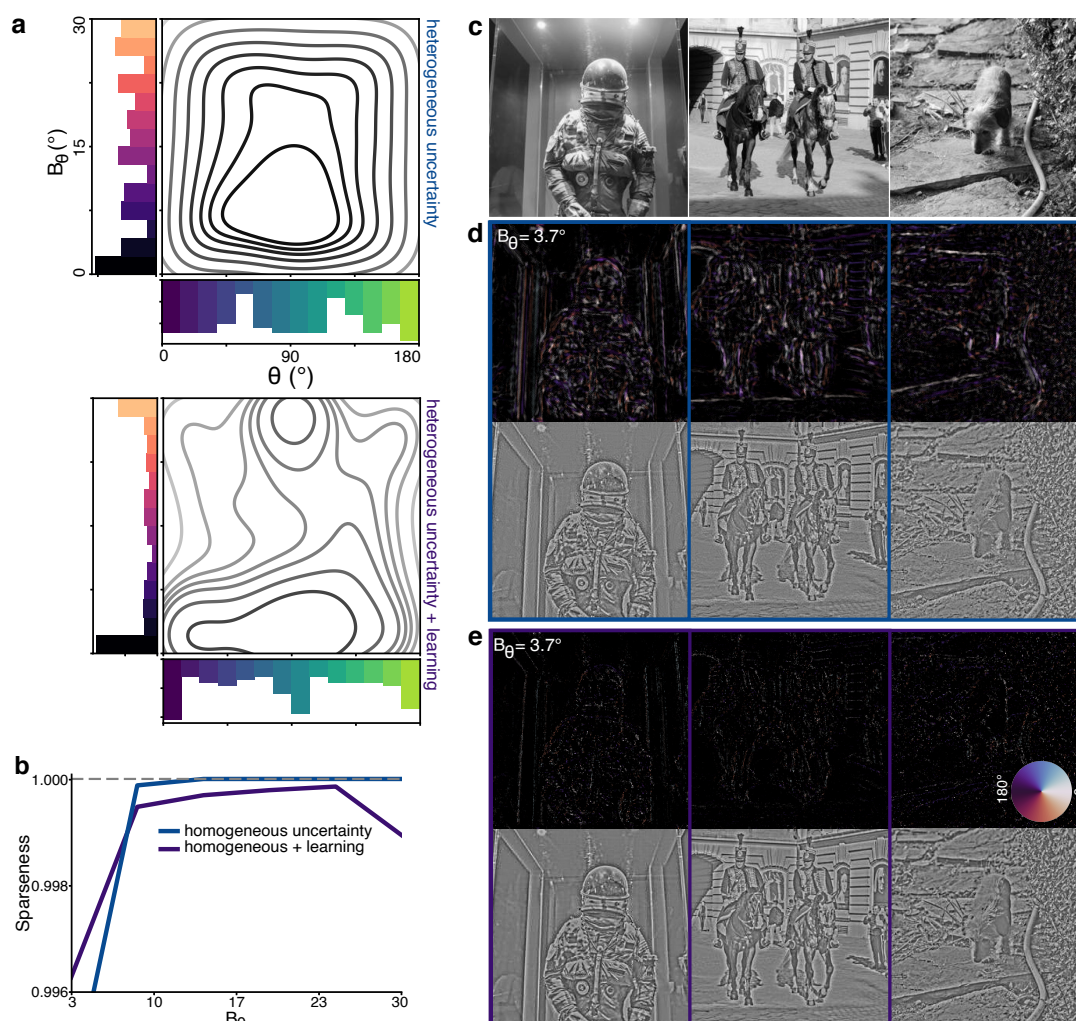
Figure 3: Learning balances coefficient distribution. **(a)** Kernel density estimation over θ and B_θ of the kernels before (top) and after (bottom) learning. **(b)** Sparseness of the dictionaries for kernel variance B_θ. Sparseness $= 1$ (i.e. no activation, as in the case of the pre-learning encoding) is represented as a gray dashed line. **(c)** Example images from the dataset. **(d)** Sparse code for high B_θ values (color coded by each coefficient's θ) and reconstructions for the pre-learned, heterogeneous variance dictionary. **(e)** Same as (d), for post-learned, heterogeneous variance dictionary. Orientation color code of the coefficients is shown on the rightmost coefficient map.

els of variance (B_θ) (Figure 2a). Specifically, learning from a dataset of natural images introduced a bias toward cardinal orientations (Figure 3a), mirroring inherent biases found in natural scenes [36], which is in contrast to the uniformly distributed initial dictionary. Furthermore, the learning process resulted in a non-uniform distribution of coefficients across multiple levels of orientation variance (Figure 3b). Notably, coefficients that were previously inactive (i.e., sparseness $= 1$) became activated at higher B_θ levels (Figure 3c-e). This led to consistent patterns in coefficient distribution across heterogeneous variance levels (Figure 3d,e). This uniformity is likely influenced by

the dataset's inherent variability. Consequently, the performance gains attributed to the learning process are contingent upon feature orientation biases (θ_0) and a redistribution of the levels of variance (B_θ), both of which should be reflective of the dataset's intrinsic structure.

3.2 Statistical properties of natural images reflect the variance of learned sparse code
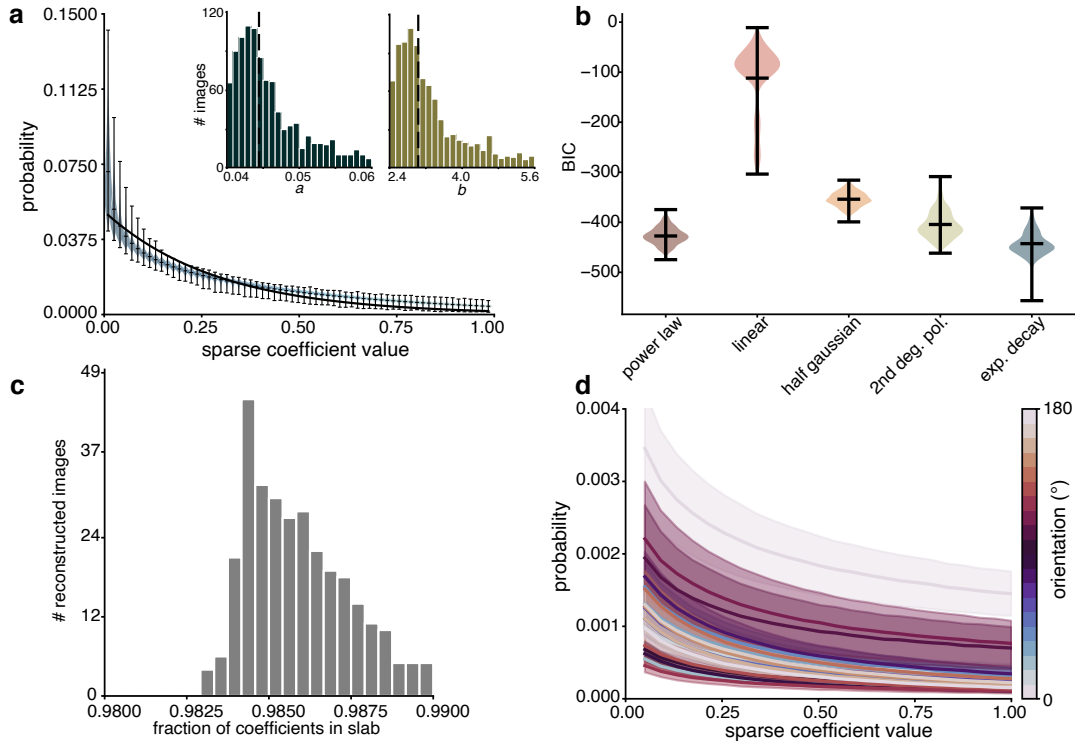
Figure 4: Spike-and-slab sparse representation of the natural images. **(a)** Distribution of the sparse coefficients values. Violin plots' central lines represent mean values, with top and bottom lines representing the extrema. For each image, this distribution was fitted with an exponential decay (black line) $y = a \cdot \exp(-b \cdot x)$, with the distributions for the parameters over the 1145 images shown in inset **(b)** Bayesian Information Criterion (BIC) for the fitting of the distribution of spikes coefficients with different alternative functions. **(c)** Proportion of zero coefficients per image, i.e., belonging to the "spike" of the distribution. **(d)** Same as (a), with coefficients split by different encoded orientation.

The criteria for the relevance of features encoded in neural networks is dictated by the statistical properties of the environment itself [9, 1]. For instance, at a fundamental representational level, the neural code for light patterns in the retina is the cumulative sum of the Gaussian distribution of luminance found in natural images [4]. At higher levels, scale distributions of visual features, in the Fourier domain, obey a $1/f^2$ power law, which once again echoes the power-law behavior of cortical responses [37, 38]. At intermediate levels, the distribution of these oriented edges can be characterized along
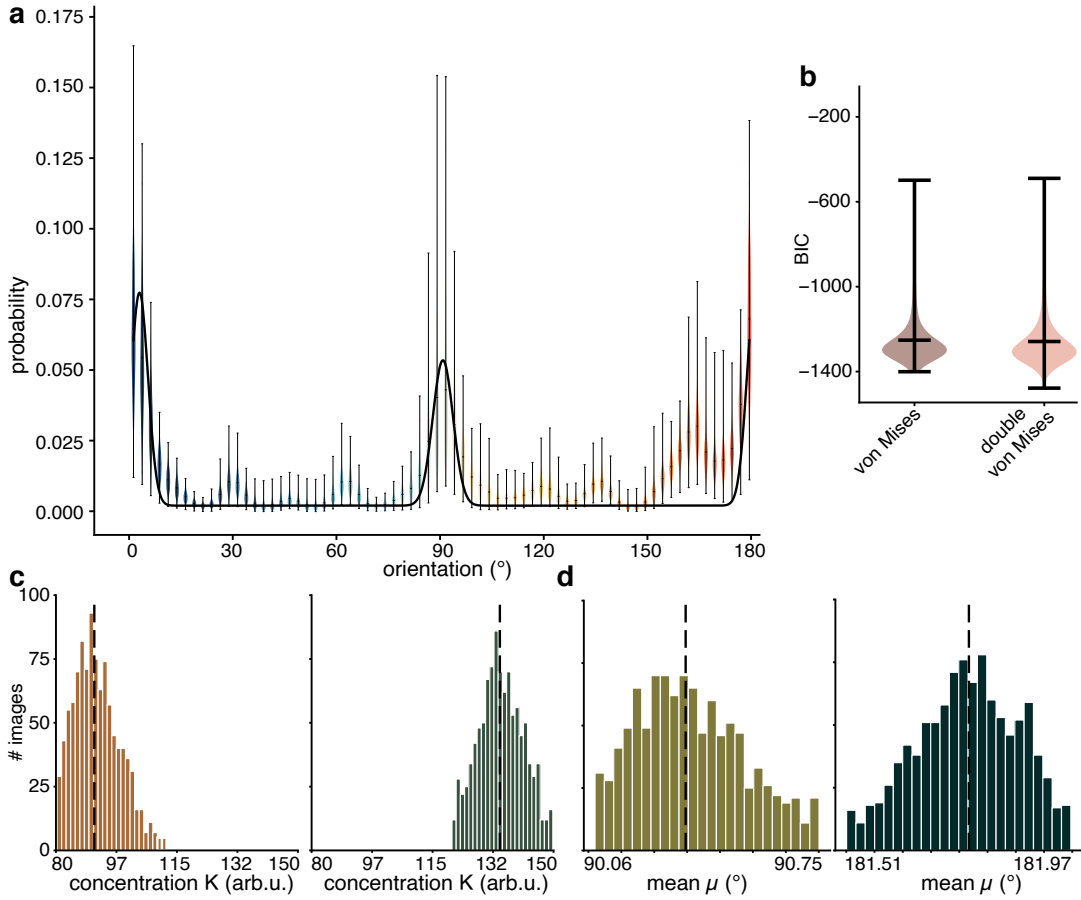
Figure 5: Orientations in natural images follow a double von Mises distribution. **(a)** Orientations of the sparse coefficients, fitted with a double von Mises distribution (black line). **(b)** Bayesian Information Criterion (BIC) for the fitting of the distribution of orientation coefficients. **(c)** Distribution of the concentration parameter κ for the first (left) and second (right) peaks of the double von Mises distribution. **(d)** Same as (c), for the mean parameter μ.

its first- and second-order moments: a median orientation, and its corresponding variance. A proper model of natural images thus depends on a proper model of both these moments, which is reflected in the response properties of primary visual cortex neurons [27]. Which of these two parameters warrants greater emphasis? Previous studies suggested that heterogeneity on both orientation and variances arises from sparse learning processes, *in silico* [2] and *in vivo* [17].

Inherently, sparse coding enforces a prior on using a minimal number of coefficients to reconstruct an image, and is thus an encoding strategy that produces a "spike and slab" distribution of activations, characterized by a predominance of zero coefficients [37] (Figure 4a-c). This imposes a prior on the representation of images at the feature-level, with a decaying exponential variation of coefficients that unfolds heterogeneously across different types of orientations (Figure 4d). Lower BIC indicate less information lost in the fitting process, and thus a better fit. Such heterogeneity in feature space stems from the fact that orientations in natural images are biased to cardinal
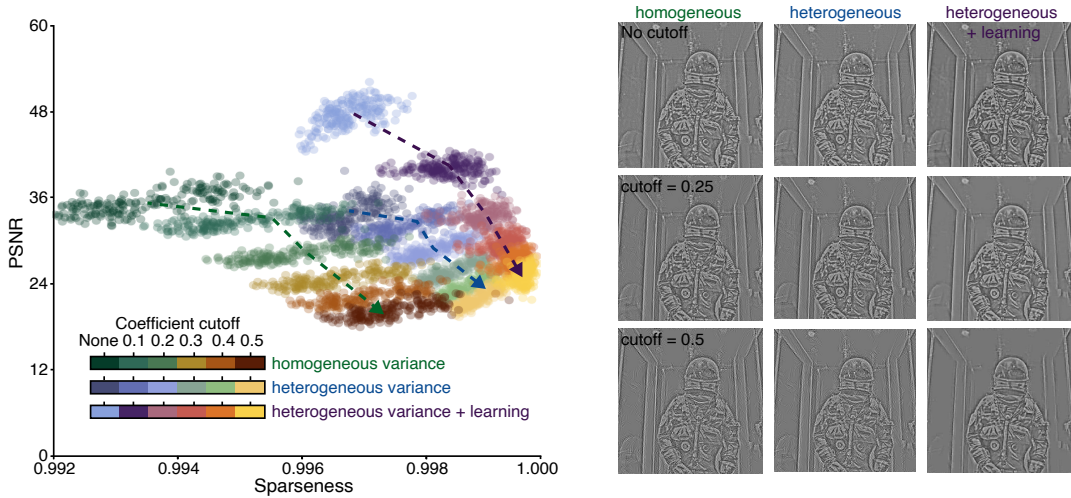
Figure 6: Sparse coefficients can be pruned for increased sparsity. **(a)** Pruning of the coefficients based on their values and resulting sparseness/PSNR for three dictionaries, with mean trajectory represented as a dashed arrow. **(b)** Reconstruction of an image with different cutoff levels.

(i.e., vertical and horizontal) orientations [39], which is echoed at the neuronal level by a cardinal bias in visual perception [40]. This biased distribution of orientation is well-captured by a double von Mises distribution in orientation space (Figure 5a,b):

$$f(x) = A_1 \exp\left(k_1 \left(\cos\left(2\pi(x - \phi_1)\right) - 1\right)\right) + A_2 \exp\left(k_2 \left(\cos\left(2\pi(x - \phi_2)\right) - 1\right)\right) \tag{11}$$

where A_1, A_2 are the amplitudes of the two von Mises distributions, k_1, k_2 are the concentration parameters for the two distributions, ϕ_1, ϕ_2 are the phase offsets for the two distributions.

This distribution is known for higher heterogeneity, and thus aleatoric variance, in natural images compared to synthetic ones [39]. At the cardinal orientations, this is also captured by the variation of the concentration parameters (Figure 5c,d) of the von Mises distributions, which underlies the notion that a proper description of natural images must be able to account for heterogeneous levels of aleatoric variance. This mandates a comparative evaluation of performance between dictionaries that emphasize a representation based on homogeneous or heterogeneous strategies, that is, emphasizing encoding mean features or their variances.

3.3   Heterogeneity improves resilience of the neural code

In addition to the previously described trade-off between performance and sparsity (Figure 2), the robustness of the representations can be further evaluated by modifying elements in the typical activation patterns. This then allows pruning less activated coefficients to further increase sparseness, testing the code's resilience to the adversarial degradation. We pruned coefficients with absolute values below a specific threshold, iterating from 0.001 to 0.5 in 6 steps. This pruning led to a construction-induced increase in sparseness, that correlated non-linearly with a decrease in PSNR for all dictionaries,
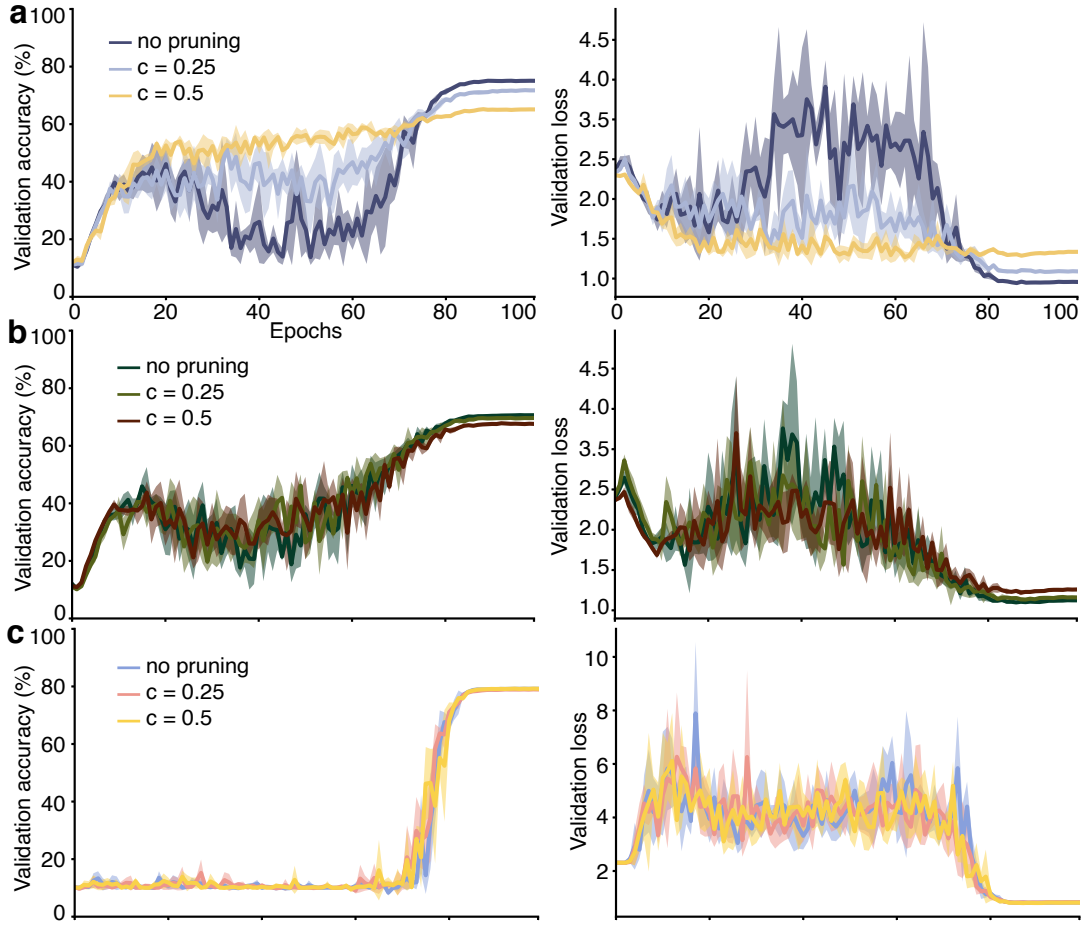
Figure 7: Deep Neural Networks (here, ResNet18), can be trained on sparse codes. **(a)** Validation accuracy (left) and losses (right) curves, for 3 different pruning levels of coefficients for the heterogeneous variance dictionary. Each network is trained across 4 random seeds, with the mean value shown as a solid line and the contour representing the standard deviation. **(b)** Same as (a), for the homogeneous variance dictionary. **(c)** Same as (a), for the heterogeneous variance dictionary, post-learning.

while maintaining interpretable representations (Figure 6), The pre-learning heterogeneous variance dictionary's PSNR demonstrated significantly greater resilience to coefficient degradation than the pre-learning homogeneous variance dictionary ($p < 0.05$ for pruning cutoff $c > 0.3$). Post-learning, both the homogeneous and heterogeneous variance dictionaries exhibited similar PSNR, reflective of their PSNR similarities before pruning (Figure 2). This emphasizes the advantage of heterogeneous variance in a dictionary, whether by construction or through learning, in bolstering resilience and efficiency for encoding natural images.

Overall, these findings show that sparse codes for natural images possess highly desirable properties when incorporating heterogeneous basis functions into a sparse model: enhanced sparseness (Figure 2d), more evenly distributed activation (Figure 3b), and increased resilience to code degradation (Figure 6a). Yet, the differences in PSNR may not necessarily translate to perceptible differences in image quality, depending on

Table 1: Mean top-1 accuracy (in $\%$) $\pm$ standard deviation across 4 random initialization of ResNet-18 for varying sparse encoding schemes of CIFAR-10. $c = 0.25$ and $c = 0.5$ indicate the pruning level of the sparse coefficients, as done in Figure 6.

Encoding scheme	No pruning	c=0.25	c=0.5
Homogeneous, pre-learning	70.65 ± 0.30	69.70 ± 0.26	67.83 ± 0.47
Homogeneous, post-learning	67.31 ± 0.20	66.24 ± 0.01	67.40 ± 0.12
Heterogeneous, pre-learning	75.08 ± 0.10	71.81 ± 0.41	65.20 ± 0.40
Heterogeneous, post-learning	$\mathbf{79.20 \pm 0.11}$	$\mathbf{78.98 \pm 0.00}$	$\mathbf{79.26 \pm 0.02}$

the context and application [41]. As such, it is necessary to investigate the potential of employing such codes in objective visual processing problems, for example, in image classification.

As a coarse analogy to a neuromorphic hierarchical sparse construction of visual processing [22, 42, 23], we trained a deep convolutional neural network to classify the sparse codes of natural images. The CIFAR-10 dataset, which was converted to grayscale in order to match the dimensionality of the dictionaries previously described, was sparse-coded and then classified using the Resnet-18 network, reaching a maximum top-1 accuracy of 79.20% in 100 epochs (Figure 7, Table 3.3). After sparse coding of the dataset, but without pruning of the coefficients, a learned dictionary initialized with a heterogeneous orientation variance basis achieved the highest classification accuracy (79.20%). This was followed by the pre-learned version of the network (75.08%), and was higher than homogeneous variance methods. Following degradation of the sparse code ($c = 0.5$), the post-learned heterogeneous variance kept similarly high performance, unlike all the other encoding scheme which showed loss of performance. The discrepancy between the deep learning performance and the previously noted similarities in PSNR and sparseness (Figure 2) underscores the significance of representing variance of low-level features in complex visual models.

Discussion

Neural systems leverage heterogeneity for increased computational efficiency [43, 44]. Here, we have explored the effects of such heterogeneous encoding of orientation variance by integrating it into a convolutional sparse coding dictionary. Our findings show that this outperforms conventional feature-representing dictionaries with fixed variance, both in sparsity and robustness, at the cost of reconstruction performance. However, these representations can be effectively employed in subsequent visual processing stages, where they result in significantly improved performances of deep convolutional neural networks. Overall, these results imply that incorporating variance in sparse coding dictionaries can substantially improve the encoding and processing of natural images.

The connection between sparse models and neural codes, which underlies the mo-

tivation behind this approach, could be further showcased using biologically plausible algorithms, such as the Locally Competitive Algorithm (LCA) [45]. Rather than enforcing sparsity through convolution as done here, this model uses a mechanism of reciprocal inhibition between each of its elements, a process that mimics particular recurrent inhibition connectivity patterns observed in the cortex [46]. This method potentially mirrors a neural adaptation of winner-takes-all algorithms, reflecting innate competition and selective activation within neural networks, and highlights the potential role of feedback loops to improve sparse coding [47]. Under this analogy, LCA could reinforce the presented framework of heterogeneity by extending it from features space (i.e., receptive fields) to also include the connectivity matrix (i.e., synaptic weights). In terms of hardware, the use of variance weighting by such a lateral inhibition mechanism could provide dynamic computational allocation for significant, unpredictable fluctuations in the data, while reducing or bypassing routine, predictable data streams. This arguably reflects the response characteristics and dynamics of cortical neurons [15, 16]. Emphasizing these pronounced shifts could streamline the data transmitted across physical channels, addressing a primary source of thermal and computational efficiency bottlenecks in neuromorphic hardware [48, 49].

In the context of image classification, our approach employing sparse coding achieved a top-1 accuracy of 79.20% on the CIFAR-10 dataset. While this falls short of the state-of-the-art performance exceeding 99.0% accuracy using color images and transformer architectures [50], it is important to note that our primary objective centered on comparing model performance with heterogeneous degree of variance in the initial layer, rather than solely pursuing state-of-the-art results. Here, the high dimensionality of the sparse-coded CIFAR-10 dataset (144 input dimensions or sparse channels), in contrast to the standard 3 dimensions in RGB images, likely contributes to this difference of accuracy. Direct integration of sparse coding with deep neural networks is a promising avenue of research that aligns with recent developments in the fields of unsupervised learning, object recognition, and face recognition. Some approaches have emphasized the ability of sparse coding to generate succinct, high-level representations of inputs, especially when applied as a pre-processing step for unsupervised learning with unlabeled data using L1-regularized optimization algorithms [51]. In several instances, the mechanism of sparse coding has been seamlessly integrated into deep networks. For instance, the Deep Sparse Coding framework [52] maintains spatial continuity between adjacent image patches, boosting performance in object recognition. Likewise, a face recognition technique combining sparse coding neural networks with softmax classifiers effectively addresses aleatoric uncertainties, including changes in lighting, expression, posture, and low-resolution scenarios [53]. Classifiers relying on sparse codes, produced by lateral inhibition in an LCA, exhibit strong resistance to adversarial attacks [54]. This resilience, potentially enhanced by heterogeneous dictionaries as explored here, offers a promising avenue for research in safety-critical applications.

The empirical evidence presented here can be interpreted as an implicit Bayesian process, wherein initial beliefs about the coefficients are updated using input images to learn the variance of visual features to represent optimally (sparse) orientations. Models with explicit integration of both model and input variance have distinct advantages in that sense. Namely, this allows to maximize model performance and minimizing decision uncertainty. In contrast, we here focused on an implicit understanding of this

relationship, demonstrating through a simple approach that vision models can benefit from factoring-in feature variance without explicit learning rules.

4 Acknowledgments

References

[1]   Eero P Simoncelli and Bruno A Olshausen. "Natural image statistics and neural representation". In: *Annual review of neuroscience* 24.1 (2001), pp. 1193–1216.

[2]   Bruno A Olshausen and David J Field. "Emergence of simple-cell receptive field properties by learning a sparse code for natural images". In: *Nature* 381.6583 (1996), pp. 607–609.

[3]   Bruno A Olshausen and David J Field. "Sparse coding with an overcomplete basis set: A strategy employed by V1?" In: *Vision research* 37.23 (1997), pp. 3311–3325.

[4]   Simon Laughlin. "A simple coding procedure enhances a neuron's information capacity". In: *Zeitschrift für Naturforschung c* 36.9-10 (1981), pp. 910–912.

[5]   Victor Boutin et al. "Sparse Deep Predictive Coding Captures Contour Integration Capabilities of the Early Visual System". In: *PLoS Computational Biology* (May 12, 2020).

[6]   Eyke Hüllermeier and Willem Waegeman. "Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods". In: *Machine Learning* 110.3 (2021), pp. 457–506.

[7]   Keisuke Nakamura and Kazuhiro Nakadai. "Robot audition based acoustic event identification using a bayesian model considering spectral and temporal uncertainties". In: *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2015, pp. 4840–4845.

[8]   Charles E Pettypiece, Melvyn A Goodale, and Jody C Culham. "Integration of haptic and visual size cues in perception and action revealed through cross-modal conflict". In: *Experimental brain research* 201.4 (2010), pp. 863–873.

[9]   Daniel L Ruderman. "The statistics of natural images". In: *Network: computation in neural systems* 5.4 (1994), p. 517.

[10]   Yann Gousseau and Jean-Michel Morel. "Are natural images of bounded variation?" In: *SIAM Journal on Mathematical Analysis* 33.3 (2001), pp. 634–648.

[11] Kaitlin L. Fair et al. "Sparse Coding Using the Locally Competitive Algorithm on the TrueNorth Neurosynaptic System". In: *Frontiers in Neuroscience* 13 (2019). ISSN: 1662-453X. URL: https://www.frontiersin.org/articles/10.3389/fnins.2019.00754 (visited on 12/21/2023).

[12] Hermann LF von Helmholtz. *Treatise on physiological optics*. 1867.

[13] Karl Friston. "A theory of cortical responses". In: *Philosophical transactions of the Royal Society B: Biological sciences* 360.1456 (2005), pp. 815–836.

[14] Gergő Orbán et al. "Neural variability and sampling-based probabilistic representations in the visual cortex". In: *Neuron* 92.2 (2016), pp. 530–543.

[15] Olivier J Hénaff et al. "Representation of visual uncertainty through neural gain variability". In: *Nature communications* 11.1 (2020), pp. 1–12.

[16] Hugo J Ladret et al. "Cortical recurrence supports resilience to sensory variance in the primary visual cortex". In: *Communications Biology* 6.1 (2023), p. 667.

[17] Robbe LT Goris, Eero P Simoncelli, and J Anthony Movshon. "Origin and function of tuning diversity in macaque visual cortex". In: *Neuron* 88.4 (2015), pp. 819–831.

[18] Honglak Lee et al. "Efficient sparse coding algorithms". In: *Advances in neural information processing systems* 19 (2006).

[19] Laurent U Perrinet. "Sparse Models for Computer Vision". In: *Biologically Inspired Computer Vision*. Ed. by Matthias Keil, Gabriel Cristóbal, and Laurent U Perrinet. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2015, pp. 319–346.

[20] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. "Atomic decomposition by basis pursuit". In: *SIAM review* 43.1 (2001), pp. 129–159.

[21] Michael Lewicki and Terrence J Sejnowski. "Coding time-varying signals using sparse, shift-invariant representations". In: *Advances in neural information processing systems* 11 (1998).

[22] Thomas Serre, Aude Oliva, and Tomaso Poggio. "A feedforward architecture accounts for rapid categorization". In: *Proceedings of the national academy of sciences* 104.15 (2007), pp. 6424–6429.

[23] Victor Boutin et al. "Pooling Strategies in V1 Can Account for the Functional and Structural Diversity across Species". In: *PLOS Computational Biology* 18.7 (2022), e1010270. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1010270. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010270 (visited on 09/14/2022).

[24] Brendt Wohlberg. "Efficient algorithms for convolutional sparse representations". In: *IEEE Transactions on Image Processing* 25.1 (2015), pp. 301–315.

[25] Brendt Wohlberg. "SPORCO: A Python package for standard and convolutional sparse representations". In: *Proceedings of the 15th Python in Science Conference, Austin, TX, USA*. 2017, pp. 1–8.

[26] Yu Wang, Wotao Yin, and Jinshan Zeng. "Global convergence of ADMM in nonconvex nonsmooth optimization". In: *Journal of Scientific Computing* 78.1 (2019), pp. 29–63.

[27] David H Hubel and Torsten N Wiesel. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". In: *The Journal of physiology* 160.1 (1962), p. 106.

[28] Sylvain Fischer et al. "Sparse Approximation of Images Inspired from the Functional Architecture of the Primary Visual Areas". In: *EURASIP Journal on Advances in Signal Processing* 2007.1 (2007), pp. 1–17. ISSN: 16876172.

[29] Sylvain Fischer et al. "Self-invertible 2D log-Gabor wavelets". In: *International Journal of Computer Vision* 75.2 (2007), pp. 231–246.

[30] Sylvain Fischer et al. "Sparse Approximation of Images Inspired from the Functional Architecture of the Primary Visual Areas". In: *EURASIP Journal on Advances in Signal Processing* 2007.1 (Dec. 2006), pp. 1–17. ISSN: 16876172. DOI: 10.1155/2007/90727. URL: http://dx.doi.org/10.1155/2007/90727.

[31] Nicholas V Swindale. "Orientation tuning curves: empirical description and estimation of parameters". In: *Biological cybernetics* 78.1 (1998), pp. 45–56.

[32] Stefan Van der Walt et al. "scikit-image: image processing in Python". In: *PeerJ* 2 (2014), e453.

[33] Hugo Ladret. "HD natural images database for sparse coding". In: *FigShare* (2023). DOI: "10.6084/m9.figshare.24167265.v1".

[34] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[35] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[36] Stuart Appelle. "Perception and discrimination as a function of stimulus orientation: the" oblique effect" in man and animals." In: *Psychological bulletin* 78.4 (1972), p. 266.

[37] David J Field. "Relations between the statistics of natural images and the response properties of cortical cells". In: *Josa a* 4.12 (1987), pp. 2379–2394.

[38] Carsen Stringer et al. "High-dimensional geometry of population responses in visual cortex". In: *Nature* 571.7765 (2019), pp. 361–365.

[39] David M Coppola et al. "The distribution of oriented contours in the real world". In: *Proceedings of the National Academy of Sciences* 95.7 (1998), pp. 4002–4006.

[40] Bruce C Hansen and Edward A Essock. "A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes". In: *Journal of vision* 4.12 (2004), pp. 5–5.

[41]   Anastasia Mozhaeva et al. "Full reference video quality assessment metric on base human visual system consistent with PSNR". In: *2021 28th Conference of Open Innovations Association (FRUCT)*. IEEE. 2021, pp. 309–315.

[42]   Martin Schrimpf et al. "Brain-score: Which artificial neural network for object recognition is most brain-like?" In: *BioRxiv* (2020), p. 407007.

[43]   Nicolas Perez-Nieves et al. "Neural heterogeneity promotes robust learning". In: *Nature communications* 12.1 (2021), p. 5791.

[44]   Matteo Di Volo and Alain Destexhe. "Optimal responsiveness and information flow in networks of heterogeneous neurons". In: *Scientific reports* 11.1 (2021), p. 17611.

[45]   Christopher J Rozell et al. "Sparse coding via thresholding and local competition in neural circuits". In: *Neural computation* 20.10 (2008), pp. 2526–2563.

[46]   Robert Coultrip, Richard Granger, and Gary Lynch. "A cortical model of winner-take-all competition via lateral inhibition". In: *Neural networks* 5.1 (1992), pp. 47–54.

[47]   Victor Boutin et al. "Effect of Top-down Connections in Hierarchical Sparse Coding". In: *Neural Computation* 32.11 (2020-02-04, November 2020), pp. 2279–2309.

[48]   Jason K Eshraghian, Xinxin Wang, and Wei D Lu. "Memristor-based binarized spiking neural networks: Challenges and applications". In: *IEEE Nanotechnology Magazine* 16.2 (2022), pp. 14–23.

[49]   Mostafa Rahimi Azghadi et al. "Complementary metal-oxide semiconductor and memristive hardware for neuromorphic computing". In: *Advanced Intelligent Systems* 2.5 (2020), p. 1900189.

[50]   Alexey Dosovitskiy et al. "An image is worth 16×16 words: Transformers for image recognition at scale". In: *arXiv preprint arXiv:2010.11929* (2020).

[51]   Raghavendran Vidya, GM Nasira, and RP Jaia Priyankka. "Sparse coding: a deep learning using unlabeled data for high-level representation". In: *2014 World Congress on Computing and Communication Technologies*. IEEE. 2014, pp. 124–127.

[52]   Yunlong He et al. "Unsupervised feature learning by deep sparse coding". In: *Proceedings of the 2014 SIAM international conference on data mining*. SIAM. 2014, pp. 902–910.

[53]   Zhuomin Zhang, Jing Li, and Renbing Zhu. "Deep neural network for face recognition based on sparse autoencoder". In: *2015 8th International Congress on Image and Signal Processing (CISP)*. IEEE. 2015, pp. 594–598.

[54]   Dylan M Paiton et al. "Selectivity and robustness of sparse coding networks". In: *Journal of vision* 20.12 (2020), pp. 10–10.

[55]   Brendt Wohlberg. "Efficient convolutional sparse coding". In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2014, pp. 7173–7177.

Appendix A - Additional Convolutional Sparse Coding details
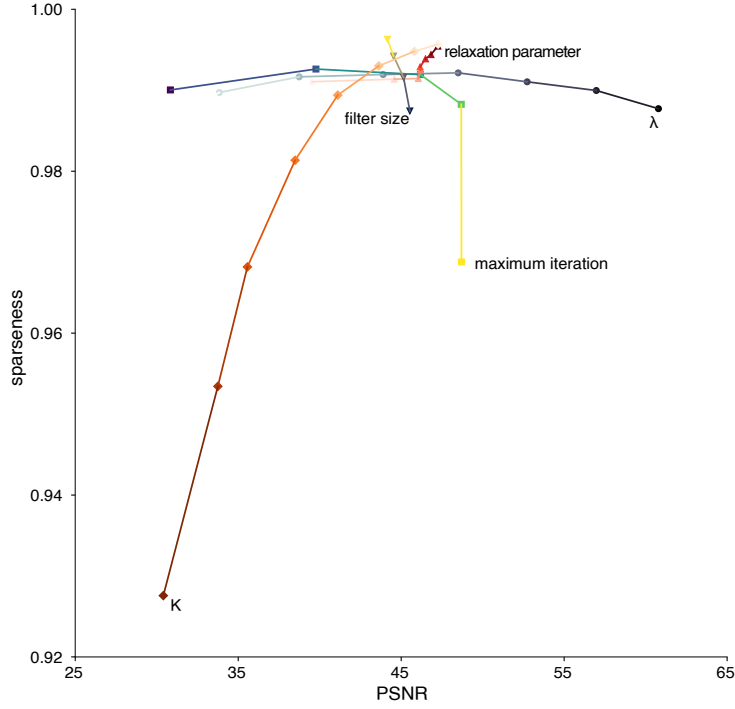
Appendix A Figure 1: Parametrization of the CSC learning algorithm. λ was varied in 8 steps in a $[0.001 : 0.1]$ range, max iteration in 5 steps in a $[10 : 1000]$ range, relaxation parameter ρ in 8 steps in a $[0.2 : 1.8]$ range, filter size in 8 steps in a $[5 : 21]$ pixels range and K in 8 steps in a $[89 : 2351]$ range.

Convolutional Sparse Coding was implemented using an Alternating Direction Method of Multipliers (ADMM) algorithm, which decomposes the problem into a standard form:

$$\underset{x,y}{\operatorname{argmin}} f(x) + g(y) \tag{12}$$

with the constraint $x = y$. This is then solved iteratively by alternating between the two sub-problems:

$$x_{i+1} = \underset{x}{\operatorname{argmin}} f(x) + \frac{\rho}{2}||x + y_i + \mathbf{u}_i||_2^2 \tag{13}$$

$$y_{i+1} = \underset{y}{\operatorname{argmin}} g(y) + \frac{\rho}{2}||x_{i+1} + y + \mathbf{u}_i||_2^2 \tag{14}$$

where ρ is a penalty parameter that controls the convergence rate of the iterations, also called the relaxation parameter. x and y are residuals whose equality is enforced by the prediction error:

$$\mathbf{u}_{i+1} = \mathbf{u}_i + x_{i+1} + y_{i+1} \tag{15}$$

ADMM can be readily applied to equation (2) by introducing an auxiliary variable Y [55], such that the problem to solve becomes:

$$\underset{\{x_k\},\{y_k\}}{\operatorname{argmin}} \frac{1}{2}||\sum_{k=1}^{K} \mathbf{d}_k * x_k - s||_2^2 + \lambda \sum_{k=1}^{K} ||y_k||_1 \text{ s.t. } \mathbf{x}_k = \mathbf{y}_k \tag{16}$$

which, following the ADMM alternation in equations (13)-(15), is solved by alternating:

$$\{x_k\}_{i+1} = \underset{\{x_k\}}{\operatorname{argmin}} \frac{1}{2}||\sum_{k=1}^{K} \mathbf{d}_k * x_k - s||_2^2 + \frac{\rho}{2}||x_k - y_{k,i} + \mathbf{u}_{k,i}||_2^2 \tag{17}$$

$$\{y_k\}_{i+1} = \underset{\{y_k\}}{\operatorname{argmin}} \lambda \sum_{k=1}^{K} ||y_k||_1 + \frac{\rho}{2}||x_{k,i+1} - y_k + \mathbf{u}_{k,i}||_2^2 \tag{18}$$

$$\mathbf{u}_{k,i+1} = \mathbf{u}_{k,i} + x_{k,i+1} - y_{k,i+1} \tag{19}$$

Appendix B - Homogeneous variance dictionary

Results from the main text are shown here for the homogeneous variance dictionary, post-learning.
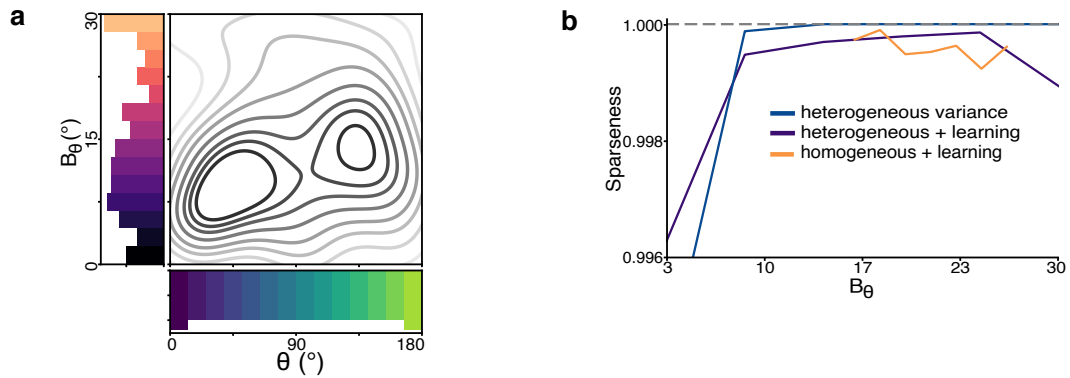
Appendix B Figure 1: Learning balances coefficient distribution. **(a)** Kernel density estimation of coefficients over θ_0 and B_θ after learning from the homogeneous variance dictionary. **(b)** Sparseness of coefficients for each B_θ. Sparseness $= 1$ is represented as a gray dashed line.
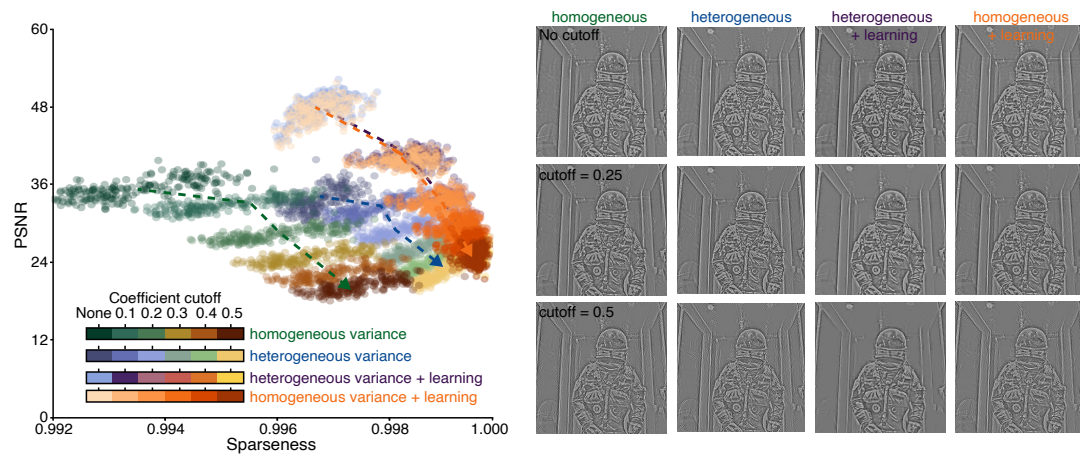
Appendix B Figure 2: Sparse coefficients can be pruned to boost sparsity. **(a)** Pruning of the coefficients based on their values and resulting sparseness/PSNR for both dictionaries. **(b)** Reconstruction of the image shown in Figure 1 with different cutoff levels.

3.5. Conclusion

Designing an efficient strategy to represent our environment is challenging, especially in a world loaded with statistical variance. Modelling such modelling visual inputs with unpredictable variance is the focal point of this thesis. In the current article, we developed a computational model that learns an optimal representation of natural images, and explored the sparseness/reconstruction trade-off of its receptive fields. By using sparse coding as a way to extract the features necessary to build a representation of natural images, we showed that a neural representation of variance is advantageous, even at the level of Deep Learning research. This ties directly to Bayesian processing, which explicitly derive mathematical rules that integrate both epistemic and aleatoric variances are advantageous (see Equation 2.3). In contrast, our study concentrates on an implicit emergent encoding of the epistemic and aleatoric variances. This also relates to the idea that computation of variance can be seen as an emergent property, which will be further emphasized in the next chapter.

A research offspring that will directly stem from this work is to train Deep Neural Networks directly on sparse coefficients that encode an optimal representation of images, rather than images themselves. The preliminary practical work on that end has already been made here, as our codebase in the article consisted of porting a sparse coding library (SPORCO [339]) into a tensor format [239], allowing for milliseconds-fast computations. Sparse representations, whether here or in the brain, are essentially distributions of binary events weighted by synaptic connectivity. As these Deep Neural Networks relying on spiking activity are set to eventually surpass regular methods [79, 119], there would be great advantage in integrating the work done here in a spiking framework. One final advantageous effect of these sparse representations is their natural property to remove noise in the input. Thus, an encoder that transforms natural images into sparse representations with variance could easily defend against noisy inputs (whether malicious or not), which would prove useful in critical applications such as medical imaging.

Aside from these machine learning considerations, the key conclusion of this article is the obvious sparseness/reconstruction trade-off involved in encoding natural images, as shown in the first figure of this article, and reproduced in Figure 3.4. This emphasizes two possible strategies to encode the distributions of orientation that makes up natural images:

— Either a neural system can use multiple orientation-tuned units to encode the full input distribution, at heavy computational and energy costs. This would be equivalent to dense sampling the likelihood $p(u|v)$ (as introduced in Equation 2.5).

— Or a neural system can use an estimate of orientation and variance, through a receptive field attuned to both of these moments of the distribution. This would then be similar to using a maximum likelihood approach (as introduced in Equation 2.6), which means finding the best surrogate function to approximate the distribution in the real input.

The fact that heterogeneous orientation-tuned functions represent the optimal trade-

off for representing natural images will serve as a natural transition towards our next chapter, which will show that V1 likely implements the first strategy for a fast first estimate, then stabilizes onto the second through recurrent connectivity.
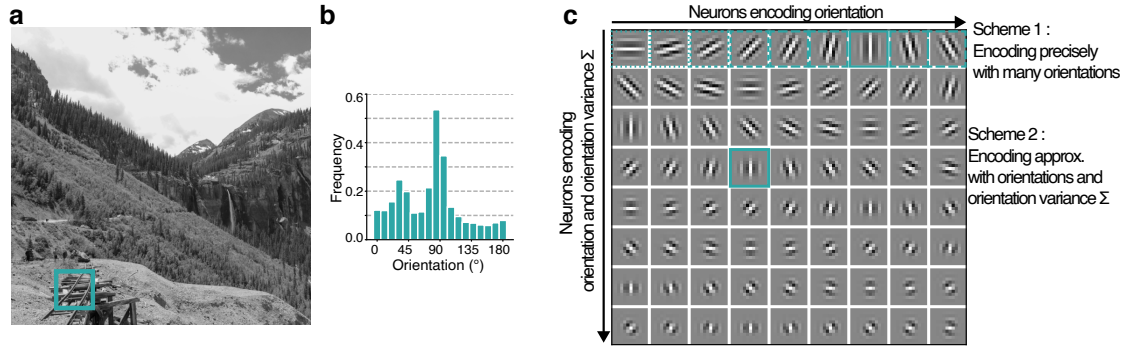
Figure 3.4. – Sparseness/reconstruction trade-off. (a) A patch of natural image, from the database generated in this article [176]. (b) Distribution of orientation from this image, extracted with Sobel filters. (c) Two possible strategies for a neural encoding of this distribution, either through costly but accurate code (Scheme 1) or sparse but less accurate code (Scheme 2).

This has been preliminarily explored by the Locally Competitive Algorithm, a neuro-inspired algorithm [261] which uses recurrent interactions between neurons to drive the sparse coding, as done in Equation 3.2, but with winner-takes-all competition:

$$\tau \frac{du_i}{dt} = -u_i + \phi_i - \lambda \sum_{j \neq i} \omega_{ij} \cdot S(u_j) \tag{3.6}$$

Where u_i is the internal state (or membrane potential) of neuron i, ϕ_i is the projection of the input onto the ith basis function (or receptive field), τ is a time constant, λ is a positive constant that scales the strength of the competition, ω_{ij} represents the degree of overlap or similarity between the receptive fields of neurons i and j, $S(u_j)$ is a function that represents the output of neuron j given its internal state u_j, typically, a threshold function. The design of this algorithm is that neurons compete with each other to represent the input, with neurons that have a dissimilar receptive field competing against one another. Using this Locally Competitive Algorithm as a model of recurrent versus feedforward interactions would also allow seeking which of these two types of connectivity creates the heterogeneous basis functions observed here. It then naturally follows that the next chapter transition from the present functional study, to pinpointing its origin in neurobiological networks.

4. Encoding of Orientation Variance through Recurrence in V1

"Representations, at a minimum,
must potentially be able to stand in for the things they represent."
Chris Eliasmith, How to Build a Brain, 2013

Summary

4.1. Introduction: Orientation Selectivity in V1

4.1.1. Orientation Selectivity and Representations in the (Visual) Cortex

As we've largely emphasized so far, orientation selectivity is a hallmark feature of the primary visual cortex. As the first functional element of the visual hierarchy, oriented receptive field form the basis of our understanding of visual processing in the cortex. This is a stereotypical role for a primary sensory cortex, which often exhibit a well-defined invariant feature code (see chapter 3) based on the sensory space they represent [45, 163], and acting as the foundation for the downstream computations. For example, the elementary units of voices is neural encoding of tonality and frequency-defined signals [25], that of bodily parts is spatially defined segments [44], and that of vision is our edges of interest, in this manuscript. Examining such foundational elements of image descriptors is a longstanding tradition in the field of visual neuroscience, especially within the framework of hierarchical networks [325].

Figure 4.1. – Hierarchical model of the visual cortex (a), oftenmost reflected in the current view of computer vision models (b). Adapted from [276].

In V1, the basis of the neural code is a visual one, and has thus the advantage of being easily conceptualized and visualized. This ease-of-understanding of V1 receptive field, supported by six decades of rich literature, makes orientation selectivity a focal point in of numerous PhD theses, serving either as a principal subject of study or, as is the case here, as an angle of attack to explore theoretical frameworks with testable hypotheses. Most often, this serves to study visual processing as a feedforward model [172], positing a sequential series of computations wherein basic visual components like edges are integrated to form progressively complex representations such as angles, textures, and eventually high-level objects, as observed in different visual cortical areas (V1, V2, V4, IT) [179]. The feedforward model proposes a structured pathway through which visual information is processed and refined at each subsequent level, contributing to a global, coherent perception of the visual world, as shown in Figure 4.1. For the sake of argumentation of this thesis, this view can be advantageous. Hierarchical modelling requires that separate features will be processed in separate areas, and as such, separate feature variances to be also localized and confined within each cortical area.

A notable limitation of the feedforward model lies in its inability to effectively distinguish between perceptions driven by bottom-up sensory input and the self-generated sensory feedback resulting from an organism's actions [333, 163]. In the case of vision, the visual flow stemming from eye movements could trigger an optokinetic reflex, akin to the reflexive response which stabilizes our view as we are reading this manuscript. Should an organism fail to discern between external and self-generated inputs, this reflex would inhibit any eye movement. As such, it follows logically that visual processes must also contain some form of feedback. A simplistic strategy to mitigate this issue could be simply cancelling the predictable consequences of self-generated sensory feedback, by sending an efference copy of a motor command. Conceptually, such transformative processes require an internal model, can be viewed as equivalents to a simulation of the external world and its consequences for the

organism. In a rather simple way (Figure 4.2), this extends the general feedforward representation framework into a predictive one, as used in this thesis [97].
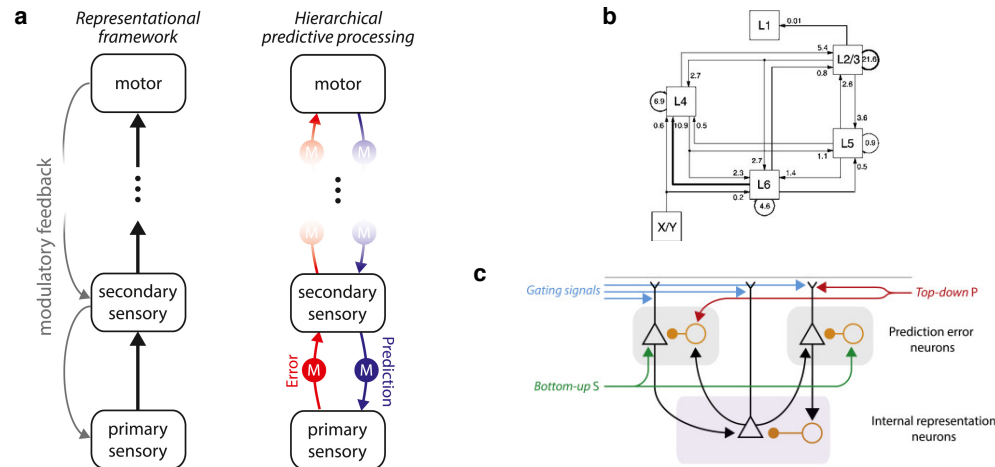
Figure 4.2. – Representation and predictive frameworks in neuroscience. (a) Predictive coding extends the classical representational framework with the addition of prediction and related error neural elements. At the relevant level of scale here, this extends the canonical microcircuit [74] (b) into a predictive neural circuit (c). Adapted from [74] and [163].

Although this doesn't challenge the notion that orientation selectivity is an optimal feature for a model of vision, it does offer an alternative perspective for the present chapter. Our aim here is to illustrate that, within a predictive modelling framework, V1 neurons don't just represent a singular feature, but also its variance, encoding a probabilistic distribution that form parts of a generative model.

4.1.2. The Origin of Orientation Selectivity in the Cortex

Pinpointing the origin of an invariant representation of images implemented within a biological neural network is a non-linear and non-trivial task. As such, there are as many theories related to "how" orientation selectivity emerges as there are theories as to "why" it exists. While it is impossible to make an exhaustive list of all phenomenological accounts and their myriad of variations, they can be grouped in three aspects:

— Orientation selectivity emerges through converging, feedforward interactions.
— Orientation selectivity is refined by local (i.e. within a cortical area) recurrent, horizontal interactions.
— Orientation selectivity is modulated by extrastriate feedback interactions.

As is the case with many complex biological questions framed as multiple choice, the answer is "all of the above, mixed non-linearly" (Figure 4.3).

The feedforward explanation for the emergence of orientation selectivity aligns intuitively with the feedforward representational framework discussed earlier. This
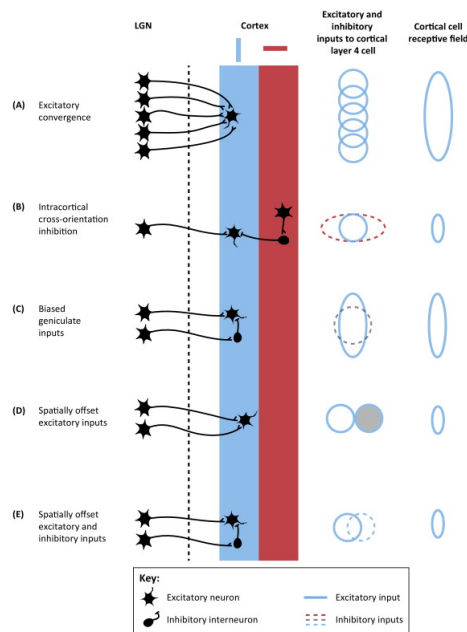
Figure 4.3. – Possible mechanisms accounting for the existence of orientation selectivity, reproduced from [330].

"canonical" model, as introduced by Hubel and Wiesel in their seminal work [138], frames orientation selectivity as arising from the convergence of isotropic receptive fields from the LGN to V1. While this has been validated experimentally several times [85, 49], such studies fail to explain the presence of (minor) orientation tuning before V1. Specifically, when cortical networks are silenced to exclusively observe LGN input to V1, the resulting input is already sharply orientation-tuned [86]. This could stem from the fact that some LGN cells are already slightly tuned to orientation, perhaps through a certain degree of retino-LGN convergence [342, 311, 326]. Indeed, even with a single thalamic oriented neuron, it is possible to obtain an excitatory orientation-tuned response in a connected V1 neuron [156].

This feedforward account of orientation selectivity is typically contrasted to the intracortical recurrent hypothesis. At the meso-scale of V1, orientation selectivity is arranged in a "map", where neighboring neurons have heterogeneous preference [120]. Long-range horizontal axons have been reported to preferentially bind to distant columns of similar orientation preferences in the cat V1, with short-range recurrent connectivity being more heterogeneous [51, 50]. This would allow having interactions between neurons tuned to different orientation at short range, as we will put forward in the present article, whilst maintaining the possibility to prime contours made of similar orientation at long range. Cross-orientation inhibition within the cortex can theoretically perform orientation sharpening [250]. This could either be responsible, on its own, for generating orientation selectivity [123], or could serve to refine the feature emerging from the feedforward convergence. Additional mechanisms encompass voltage-sensitive mechanisms with precise location within the dendritic tree, but also
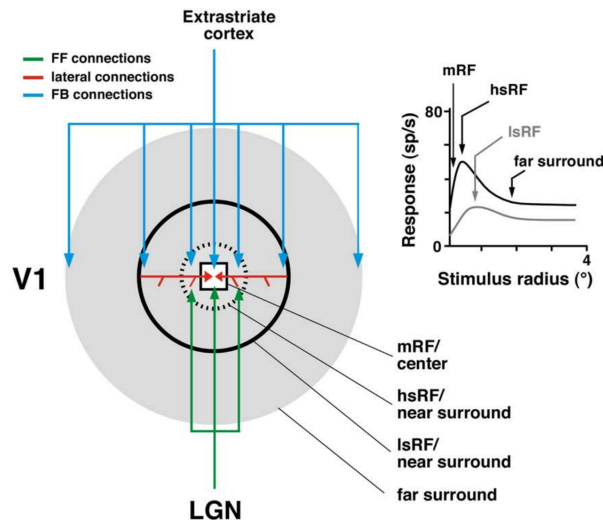
Figure 4.4. – Spatial extent and response induced by feedforward (FF), lateral and feedback (FB) connections. Adapted from [11].

intracortical excitation emanating from cells that are tuned to analogous orientations [329, 250, 330].

This also yields an interesting counter-observation, given that precisely organized orientation maps are solely present in V1 of carnivora and primates [144]. Rodent, on the other end, have a "salt-and-pepper" (random) topology of orientation detectors, with no specific spatial mapping onto the cortex. However, through a delicate balance of excitation and inhibition, it is also possible that these networks recurrently create orientation selective neural activity [123]. This is corroborated by the notion that recurrent interactions among potentially isotropic cortical neurons can yield properties akin to those arising from feedforward interactions. In other terms, this implies that functional convergence, whether feedforward or recurrent, is an inherently viable mechanism for the emergence of orientation selectivity. Extending this idea, one can also think of the complex cells (see Introduction 2.1.1.2.) as either a convergence of simple cells, but also as recurrently "amplified" simple cells [48].

Given that these rodents also exhibit orientation tuning within the LGN [311], this raises additional questions. Do these phenomena reflect unique characteristics specific to certain species, or do they suggest a more comprehensive need to re-evaluate established beliefs about orientation selectivity? This question becomes especially pertinent when we consider the wide range of species exhibiting pronounced orientation tuning within the cortex. This possible difference in strategy for orientation selectivity might also speak of different strategies for processing associated variance, between primates, cats and mice. As we shall see at the end of this chapter, primates and cats seemingly exhibit similar behavior, but recordings on mice are currently underway in our laboratory, and will be discussed in the Conclusion of this thesis.

Finally, V1 receives substantial and potent feedback from extrastriate cortical areas, which plays a pivotal role in shaping orientation selectivity. A salient illustration of
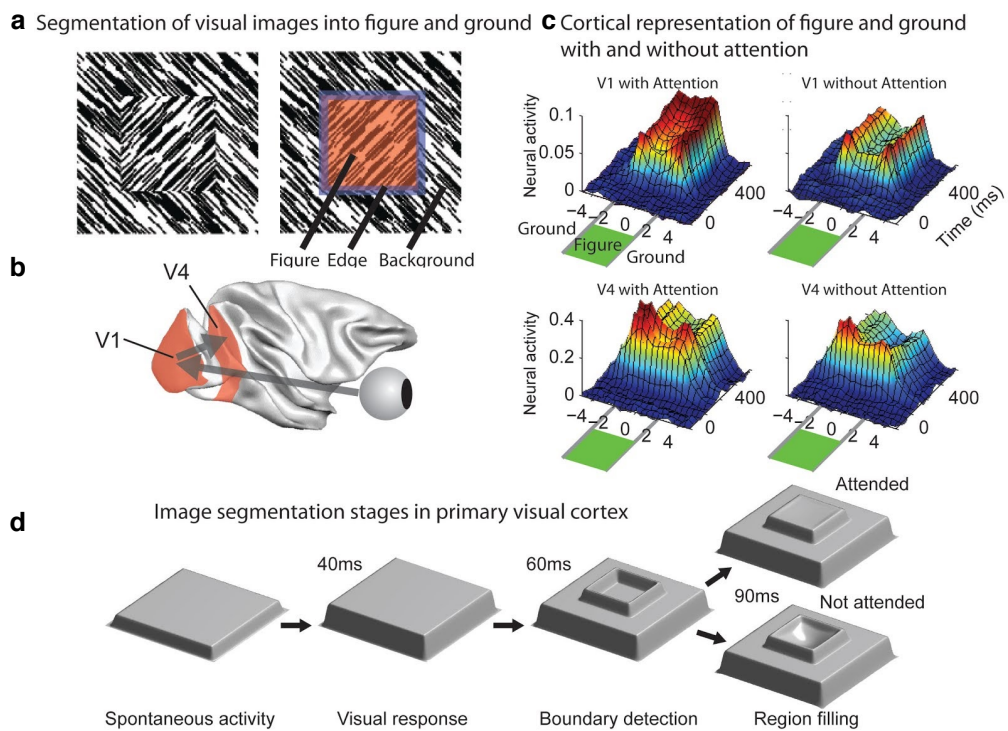
Figure 4.5. – Feedback from extrastriate areas modulates V1 orientation selectivity. (a) In a task of segmenting various elements in a naturalistic context, (b) both orientation selectivity and texture selectivity, from V1 and V4 respectively, are involved. (c) The receptive field of V1 neurons changes with attention towards the figure, which is based on feedback from V4, as summarized in (d). Figures reproduced from [248]

this intricate relationship is observed in the task of figure-ground segmentation: the feedback from extrastriate areas (here, V4), allows V1 to not only to delineate the edges with precision, but also to a fully-fledged figure in meticulous details [248] (Figure 4.5). This feedback exhibits a broader spatial extent compared to the feedforward receptive field, adding a layer of complexity to our understanding [11].

In predictive coding terms, such feedback mechanisms are modeled as carrying the predictions from higher-order regions to V1. It is often implied by predictive coding these feedbacks should be modulatory only [20], in order to send prediction from higher- to lower-order areas that can suppress prediction errors produced by bottom-up input. This is however neither true [220, 341, 65] nor required, as mixed excitatory/inhibitory influence can mediate the construction of prediction and prediction errors locally. The modular nature of this activity will be discussed further in the next chapter.

Overall, one can see how orientation selectivity in V1 is a well-defined invariant representation of low-level features of the world. As such, it serves as the perfect predictions of these features, and hence, is constrained to variance weighting, as developed in Equation 2.28. While there is no clear consensus on the origin of orientation selectivity in V1, there is no debate that the cortical circuitry is dedicated to maintain it, through a complex mix of many neural activities, that must be first recorded and then deciphered to the best of the experimenter's ability.

4.2. Methods: Visual Electrophysiology and Neural Decoding

4.2.1. Recordings Tools of the Brain
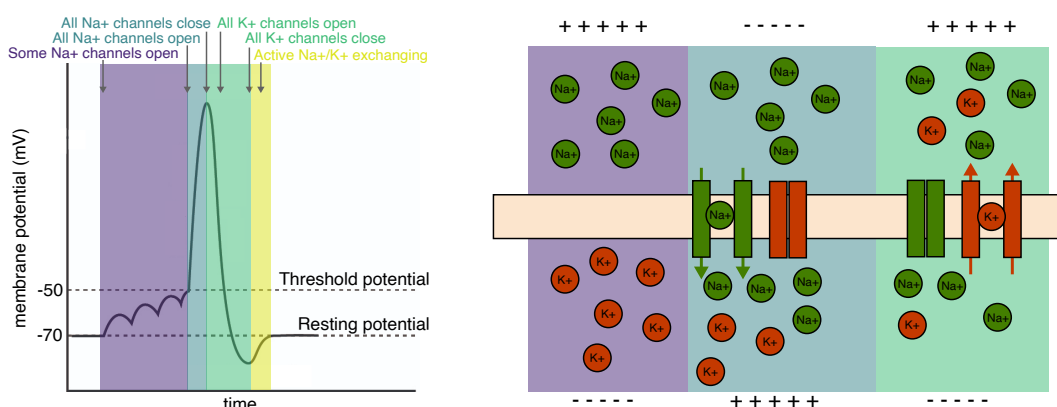
Figure 4.6. – Schematic illustration of the action potential on the membrane potential (left) and its respective ionic currents (right). The final active Na+/K+ exchanging is not illustrated.

Neurons, as the functional units of the nervous system, communicate through

sophisticated electrochemical activity, which involve using the flow of ions to generate electrical gradients [154]. To simplify this process, neurons effectively maintain an electrochemical gradient that puts them at an electric potential of -70mV with respect to their local environment. Upon binding of a neurotransmitter, channels specifically let Na+ ions flow through, depolarizing the neurons up to approximately $+40$mV. The neuron then activates energy-consuming K+ channels, re-establishing a polarized electrochemical gradient, with a slight overshooting. This change of activity propagates from the neuron's soma to its axon terminal, where it induces the fusion of synaptic vesicles through the entry of Ca2+ ions, resulting in the release of neurotransmitters into the synaptic cleft, which in turn can alter the conductance of the post-synaptic neuron (Figure 4.6). The modulation of conductance in the post-synaptic neuron is crucial as it may lead to the generation of a new action potential, thus perpetuating the chain of neural communication.
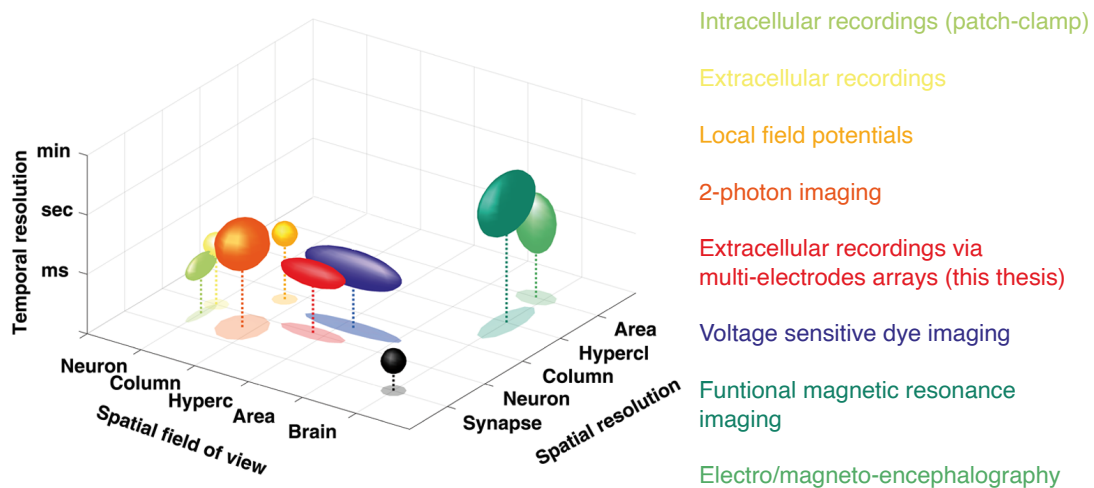
Figure 4.7. – Spatial field of view, and spatio-temporal resolution of various methods of recordings of the nervous system, from [52].

This mechanism, (over)simplified here for clarity's sake, forms the basis of neural communication. Thus, all methods employed for recording brain activity are inherently also based on it. The lowest possible level (in terms of spatial and temporal resolution) of such recording methods consists in approaching a glass pipette with an electrode to the membrane of the neuron, then forming a seal to directly record that patch of membrane [265]. This allows to record single ionic channels, which is highly effective for mapping conductance of ion to single molecules. By breaking this seal, it is possible to record the whole neuron, but also control its dynamics using current or voltage injection. Whether at the single channel or membrane level, such methods are known as "patch-clamp" [224], either in "current-clamp" or "voltage-clamp" mode.

Patch-clamp techniques, while allowing to measure single neuron activity in exquisite details, cannot record from populations of neurons. Consequently, for questions related to the internal representations of certain features with an ensemble of neural activity, it is often best to measure electrical potential variations outside the neuron,
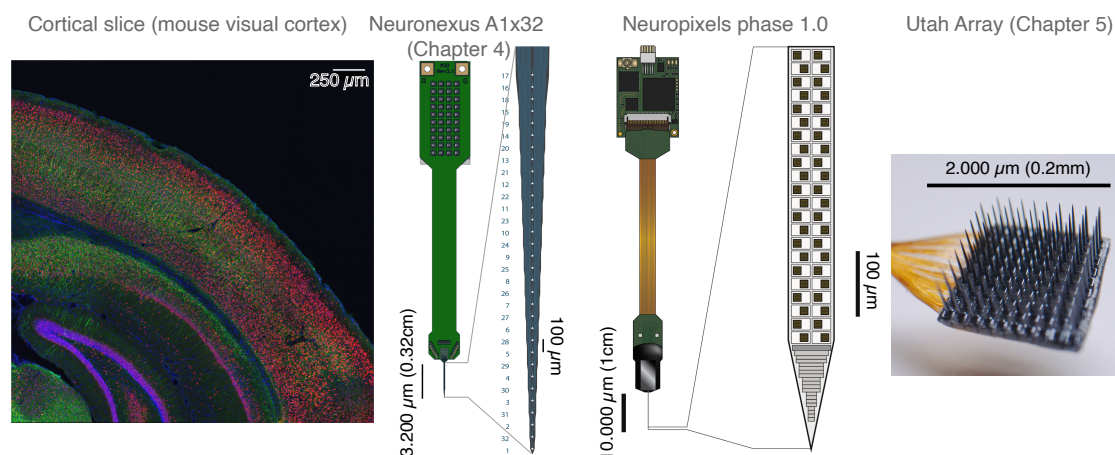
Figure 4.8. – Various types of extracellular electrodes, with cortical slice of mouse visual cortex (acquired by Geneviève Cyr) for size reference.

which does not involve precisely putting an electrode on the neuronal membrane [43]. Rather, this method necessitates the insertion of electrodes into the brain, a procedure historically done using tungsten electrodes [137, 105], but nowadays predominantly achieved using multi-channel electrodes made from complex alloys [68]. This method of recording is the one used throughout this thesis (Figure 4.8).

Multi-channel electrodes permit the recording of electrical activity from a multitude of neurons simultaneously, allowing extensive data collection and nuanced understanding of neural dynamics [296, 295]. As of 2023, the advanced state of this technology allows for the recording of approximately one thousand neurons concurrently. Presently, we employed electrodes with lower channel counts, specifically 32 to record neural activity. The spatial configuration of these electrodes is varied, but typically that of a linear probe, with contacts all along the vertical axis of the probe. In chapter 5, we will use a grid-like variation of that probe, which does not allow to probe for layer-specific computations, but for horizontally distributed ones instead. Signals recorded by extracellular electrode are a mixture of activity from multiple currents and neurons near each electrode site. To assign each extracellular event to a given neuron, several algorithms exist, each performing a different (but related) version of a "spike-sorting" process. Here, Kilosort3 [236, 260] is used, which performs template-matching based on the extracellular events' waveforms, and has been shown to achieve state-of-the-art performance for multi-channel electrodes. Nonetheless, its output requires a (lengthy) post-processing step by the experimenter, to merge and dissociate entangled neuronal activity, which is here done with a graphical interface called "Phy" [260].

Depending on the frequency at which the signal is filtered, one can also record the local field potential from these electrodes. This constitutes a non-assigned, summed signal of all neuronal activity from neighboring units [152, 131]. This technique is quite useful, because by recording the polarities of the events, one can observe the "sink" of neural activity arriving after a visual a stimulation in the layer IV of V1, and

the corresponding "source" in deeper and upper layers [268]. These sink/source terms refer to the influx of Na2+ ions (creating a "sink" outside the neurons) after a depolarization, and inversely for the "source". This technique is used in the present article to assign a cortical layer to the recorded neurons [178].

For the sake of a comprehensive overview, let us continue exploring the various recording tools of the brain. As far as neuron counts is concerned, increases in scale then require a change from electrodes to optical recording methods. A notable example that drives many experiments at the time of writing is calcium imaging, a technique driven by the principle that $Ca2+$ influx related to an action potential [302, 91]. This method enables the simultaneous recording of the activities of tens of thousands of neurons [303]. However, its application is hampered by depth limitations, owing to the inherent constraints of optical tools in penetrating the depth of the cortex. To overcome this limitation, the exploitation of advanced physical phenomena such as two- or three-photon effects can be employed, allowing for increased depth of imaging [56]. However, this entails a surge in the complexity of the acquisition system. The implementation of this technique also necessitates unobstructed optical access to the cortex, involving the removal of the skull for acute experiments and, in some cases, its substitution with viewing glass for chronic studies. The temporal resolution of the calcium imaging is on the scale of tens of milliseconds, which can hide single-neuron dynamics. To that end, there exists the possibility of loading the cortex with Voltage-Sensitive Dye (VSD), which loads the cortical areas with a molecule sensitive to changes in voltage [52]. This technique grants superior temporal resolution [53], allowing for detailed imaging across the entire cortical area, but most importantly, sub-action potential recordings, because it does not rely on the Ca2+ influx of the action potential.

At the culmination of the spectrum of recording methodologies are macroscale recordings, which encompass a range of diverse techniques, each providing unique and non-invasive insights into brain activity. Direct measurements of electrical activity can be obtained through the widely known Electroencephalography (EEG) [210], which measures the sink/source currents of the large cortical pyramidal neurons, offering insights into the global electrical activity of the brain [168]. This is, in a way, the extra-cranial version of the recordings of Local Field Potentials described above. For finer resolution, one can rely on Magnetoencephalography, a sophisticated method that records the orthogonal magnetic field corresponding to the electrical signals generated by neural activity [125]. The acquisition setup is however significantly more complex, necessitating the integration of supra-conductive sensors to precisely capture the delicate magnetic fields associated with neural currents. Lastly, the mainstream method for large-brain analysis is functional Magnetic Resonance Imaging, a technique that detects the magnetic signatures of oxygen-binding molecules [340, 193]. These molecules flow into specific regions of the brain in response to neural activity to restore the local energy supply to neurons [246]. By mapping these blood flow changes, this technique enables the indirect observation of neural activity, providing (rough) insights into the functional organization of the brain

Each of these techniques, despite their complexities and inherent limitations, con-

tributes uniquely to our expansive view of neural activity, providing distinct insights into the dynamic interplay of neural circuits. These advanced methodologies, in conjunction with continuous advancements in neuroscientific research, are instrumental in unraveling the intricate tapestry of neuronal interactions and communications.

4.2.2. Making Sense of the Recordings

It may appear straightforward to correlate a single visual input with the quantity of spikes discharged by a single neuron and call it a code. However, this seemingly straightforward approach, formally called "rate coding", involves the prior that the frequency of spike emissions by a neuron is presumed to signify its sensitivity to a particular stimulus. Under this paradigm, a higher spike rate is interpreted as indicative of greater neuronal responsiveness to the stimulus. As we have detailed in the introduction, this assumption stands in contrast to substantial evidence suggesting that neurons inherently strive to minimize their energy consumption as much as possible. The rate-based model, although used by Hubel and Wiesel in the original orientation-selectivity articles [137, 138], does not capture the full breadth of possible information strategies contained in neural recordings.

Navigating through all the intricate mechanisms of neural coding could mandate an extra dedicated manuscript. However, the main alternative hypothesis to rate coding is temporal coding [78], which posits that neurons encode information through intricate patterns in the timing of their spikes, implicating diverse aspects such as delays [118], synchronicity, and first-spike times in the encoding process. These multifaceted strategies all imply a different way to look at the data. This can be manageable for recordings with a low neuron count, such as an implanted electrode in a given area for an extended period of time [228]. With great sampling power comes little understandability, and with experiments with hundreds of neurons, it becomes harder to disentangle the meaning from the noise in the recordings. This is where machine learning-based approaches come in, allowing the experimenter to study its recordings with minimal prior assumptions.
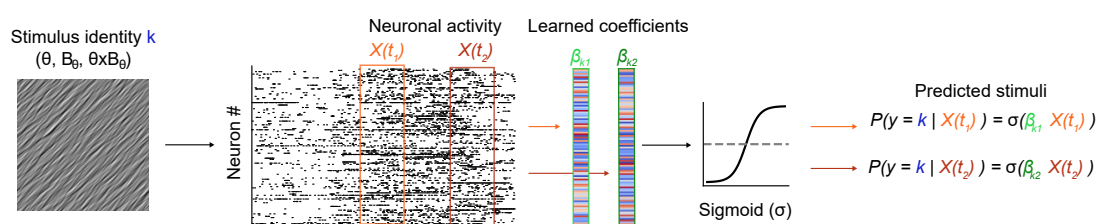
Figure 4.9. – Illustration of logistic regression applied to neuronal decoding. Additional details on the mathematical framework can be found in the chapter's article.

As the sensory organs can essentially be seen as "encoding" the features of the world, using machine learning (or really, any non-computationally trivial) techniques to understand the neural activity is referred to as "neural decoding" [208, 113]. Numerous

approaches exist for neural decoding, mostly aiming to categorize neural activity into distinct types of stimuli through labelled, supervised training. The efficacy of a decoding algorithm is contingent upon its ability to discriminate between different types of neural activity; the more refined the discrimination, the more accurate the encoding of the feature to be classified within the neuron. Intuitively, this is akin to having a better separate of two conditional distributions in the neural substrate, which yields better classifier performance as the boundary between these distributions becomes better defined. For a given algorithm with optimal parametrization, improvement of the classification based on two different recordings means that features sought-after in the recordings have become more disentangled by the neurons in feature space [29].

This forms the basis of the present article. In this research, we employ logistic regression as our neural decoder, a method praised for its simplicity, efficiency, and biological plausibility. While the implementation details are elaborated within the article, the foundational principle is depicted in the accompanying figure above. Logistic regression serves as a robust tool to unravel the complex patterns within neural datasets, allowing for the elucidation of the nuanced interactions and responses of neurons to varied stimuli.

4.3. Article: "Cortical Recurrence supports Resilience to Sensory Variance in the Primary Visual Cortex"

The following article represents a key contribution of this thesis, notably by introducing two novel neuronal responses that contribute to sensory variance encoding. Using a computational model, we validate these findings through a computational model of intracortical connectivity, which serves as the cornerstone for our argument as to how the brain processes distributions of naturalistic inputs in subsequent chapters.

Full citation is as follows: Hugo J Ladret, Nelson Cortes, Lamyae Ikan, et al. "Cortical recurrence supports resilience to sensory variance in the primary visual cortex". In: *Communications Biology* 6.1 (2023), p. 667

communications
biology

ARTICLE

OPEN

Check for updates

Cortical recurrence supports resilience to sensory variance in the primary visual cortex

Hugo J. Ladret [1,2 ✉], Nelson Cortes[2], Lamyae Ikan[2], Frédéric Chavane [1], Christian Casanova [2] & Laurent U. Perrinet [1]

Our daily endeavors occur in a complex visual environment, whose intrinsic variability challenges the way we integrate information to make decisions. By processing myriads of parallel sensory inputs, our brain is theoretically able to compute the variance of its environment, a cue known to guide our behavior. Yet, the neurobiological and computational basis of such variance computations are still poorly understood. Here, we quantify the dynamics of sensory variance modulations of cat primary visual cortex neurons. We report two archetypal neuronal responses, one of which is resilient to changes in variance and co-encodes the sensory feature and its variance, improving the population encoding of orientation. The existence of these variance-specific responses can be accounted for by a model of intra-cortical recurrent connectivity. We thus propose that local recurrent circuits process uncertainty as a generic computation, advancing our understanding of how the brain handles naturalistic inputs.

[1] Institut de Neurosciences de la Timone, UMR 7289, CNRS and Aix-Marseille Université, Marseille, France. [2] School of Optometry, Université de Montréal, Montréal, Canada. ✉email: hugo.ladret@univ-amu.fr

ARTICLE

Selectivity to the orientation of visual stimuli is an archetypal feature of the neurons in the mammalian primary visual cortex (V1)[1], which has been historically studied using low-complexity stimuli such as oriented gratings[2]. While this approach offers a clear hypothesis as to what neurons are responding to, it only probes for neural selectivity to individual input parameters, such as orientation or spatial frequency. Natural vision, however, involves rich cortical dynamics[3] integrating a mixture of multiple local parameters and global contextual information[4]. Hence, a majority of our understanding of V1 relies on neural responses to single inputs in orientation space, rather than naturalistic responses to multiple orientations.

This knowledge gap is not trivial, as the variance of distributions of sensory inputs is a fundamental cue on which our brain relies to produce coherent integration of sensory inputs and prior knowledge of the world[5,6] in order to drive behavior[7]. According to Bayesian inference rules, low-variance inputs are processed through fast feedforward pathways, whereas higher sensory variance elicits a slower, recurrent integration[8]. How the brain performs computations on variance is not yet fully understood. In V1, it has been shown that single neurons undergo nonlinear tuning modulations as a function of their input's variance[9] which can serve as a functional encoding scheme[10,11]. These recent results align with earlier models of recurrent cortical activity of V1[12,13] and also match psychophysical measurements in humans[14–16]. While it seems that local interactions within V1 are sufficient to encode orientation variance[17], the quantification of single neuron responses, their dynamics and their link to a functional population encoding of variance remains to be established.

Here, we investigate the neural basis of variance processes in V1 using stimuli matching the orientation content of natural images[18]. We present a quantitative analysis of single neurons' variance-tuning functions, as well as their dynamics, reporting heterogeneous modulations. Two archetypal response types emerge in V1, one of which relies on predominantly supra-granular neurons that maintain robust orientation tuning despite high sensory variance, allowing them to co-encode orientation and variance, and enhancing V1's orientation distribution encoding. A well-established V1 intracortical recurrence model accounts for these resilient neurons, aligning with canonical Bayesian frameworks[6] and suggesting uncertainty computations as a new generic function for local recurrent cortical connectivity.

Results

Single-neuron response in V1 depends on input variance. We recorded neural activity from 249 anesthetized cat V1 neurons and measured orientation-selective responses to naturalistic images called Motion Clouds[18]. These stimuli are band-pass filtered white noise textures and offer three advantages over both simple grating-like stimuli and complex natural images. First, they enable fine control of mean θ and variance, controlled by B_θ, of orientation distributions through a generative model, thereby reproducing natural images' oriented content (Fig. 1). Second, as they are stationary in the spatial domain, they only probe orientation space, excluding any second-order information exploitable by the visual cortex[19]. Third, by conforming to natural images' $1/f^2$ power spectrum distribution[20], they attain a desirable balance between controllability and naturalness[21]. We generated 96 Motion Clouds by varying mean orientation θ between 0° and 180° in 12 even steps and variance B_θ between ≈0° and 35° in eight evenly spaced steps.

All recorded neurons displayed orientation selectivity to Motion Clouds. Nearly all (98.8%, $p < 0.05$, Wilcoxon signed-rank test) units maintained their preferred orientation when variance B_θ increased, while the peak amplitude of the tuning curve diminished significantly (95.1% units, $p < 0.05$, Wilcoxon signed-rank test, 73.1% mean amplitude decrease for $B_\theta = 35$°). Only 28.5% of the recorded units were still tuned for $B_\theta = 35.0$° stimuli ($p < 0.05$, Wilcoxon signed-rank test). Thus, increasing input variance reduces single neuron tuning, which manifests heterogeneously across neurons, as evidenced by two representative single units shown in Fig. 2a. Neuron A illustrates single units which are no longer orientation-tuned when variance B_θ reaches 35° ($W = 171.0, p = 0.24$, Wilcoxon signed-rank test), unlike neuron B ($W = 22.5, p = 10^{-6}$) which exemplifies the aforementioned 28.5% variance-resilient units. These response types are characterized by functions relating B_θ to the goodness of tuning (circular variance, CV), named here variance-tuning functions (VTF, Fig. 2b). Such VTFs represent the input/output transformation in variance space, and are well-fitted with Naka-Rushton functions[22] (Supplementary Fig. 2a). This allows to summarize variance modulations using only three parameters: n, the VTF non-linearity; $B_{\theta 50}$, the input variance level for the tuned-untuned state transition; and f_0, the orientation tuning goodness for lowest-variance inputs. Overall, VTFs exposed diverse responses to variance among V1 neurons, with median values outlining a characteristic VTF that is slightly nonlinear, with a changepoint at $B_\theta = 19.2$° (Fig. 2c). In other words, most neurons tend to change abruptly in tuning when input variance reaches 19.2°, after which the response becomes less sensitive to orientation. Alternative metrics were also calculated, including variance-half width at half height (HWHH) and variance-maximum response functions (Supplementary Fig. 2b–e). Although HWHH displayed patterns resembling VTFs, we elected to not use it, as its reliance on fits, its consequent susceptibility to fitting artifacts, and its similarity with CV are not desirable properties. Since CV also inherently accounts for the firing rate at the preferred orientation (see "Methods"), we relied on this metric to describe both maximum amplitude and goodness of tuning in a single metric.

Orientation variance impacts not only orientation tuning but also the dynamics of the response of V1 neurons (Fig. 3). Interestingly, both effects are linked, as demonstrated by the two example VTFs: neuron B, which exhibited orientation-tuned responses for $B_\theta = 35$° inputs (Fig. 2a), also had a slower time-dependent change of goodness of tuning (relative min. of reduction of 42% of max. CV at 200 ms post-stimulation onset, $B_\theta = 0$°) compared to neuron A (relative min. of 26% of max. CV at 90 ms post-stimulation onset, Fig. 3b). These dynamical modulations were also heterogeneously distributed among the population, significantly more spikes emitted 200 ms after stimulation onset for $B_\theta = 35$° (Fig. 3d, $U = 14936.0, p < 0.001$, Mann–Whitney U-test). In summary, orientation variance induces changes in both tuning and dynamics of V1 neurons, revealing two archetypal types of response: either fast in time and nonlinear with respect to variance (neuron A) or slow in time and linear with respect to variance (neuron B).

Multiple types of variance responses are found in V1. To properly characterize the two aforementioned types of responses to variance, we separated the recorded neurons into two groups using K-means clustering the Principal Components (PC, Fig. 4) of the neuronal responses. Clustering was performed on the VTFs (Fig. 4b), tuning statistical measurements (Fig. 4c, d) and response dynamics (Fig. 4e, f). We used the first 2 PC for clustering the data, which accounted for 39.1% of the cumulative variance (Supplementary Fig. 4a), and chose two clusters based on the number of example responses and the empirical absence of an elbow[23] in the Within-Clusters-Sum-of-Squares (WCSS) curve (Supplementary Fig. 4b). This splits the data into a cluster of 164

ARTICLE
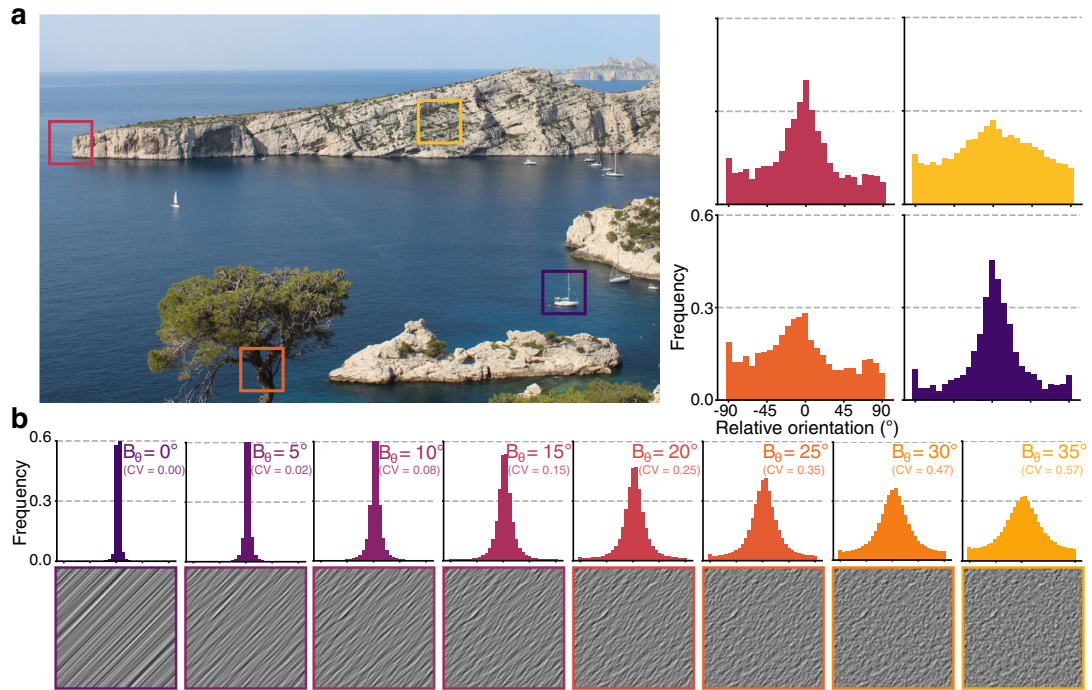
Fig. 1 Variance of orientation distributions characterizes local regions of natural images. a Distributions of orientation from four 200x200 px regions of a natural image (picture taken by H.J.L.) obtained by a histogram of oriented gradients (32x32 px/cell), centered around the most frequent orientation. **b** Motion Clouds, naturalistic stimuli (bottom row) with mean orientation $\theta = 45°$ and increasing variance (B_θ) from left to right. Distributions of the orientation of the stimuli are shown on the upper row. Circular Variance (CV) of the distribution is shown for comparison.
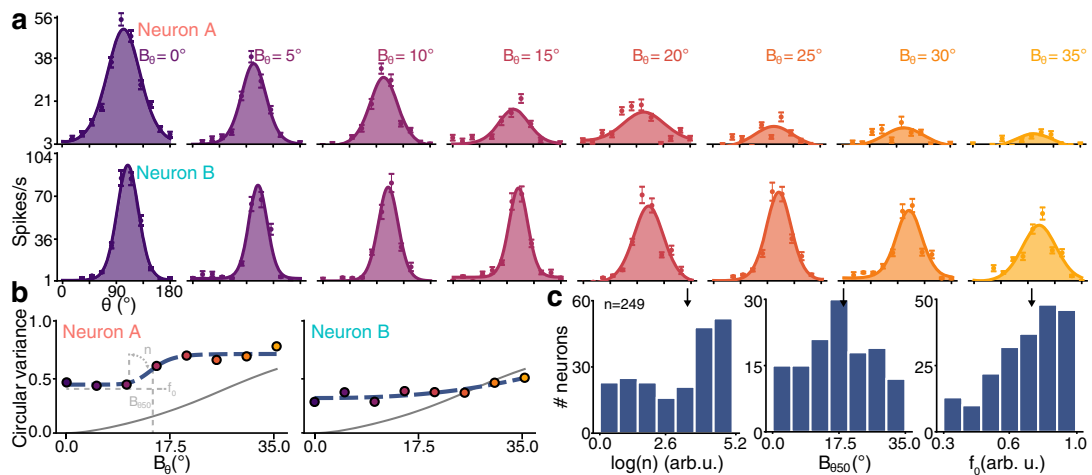
Fig. 2 Single-neuron tuning correlates with input variance. Additional examples are shown in Supplementary Fig. 1. **a** Tuning curves of two neurons responding to Motion Clouds of increasing variance B_θ. Dots indicate the mean firing rate across trials (baseline subtracted, 300 ms average), error bars are the standard error and lines represent a fitted von Mises function. **b** Variance-tuning functions (VTF), relating the change of orientation tuning measured by the circular variance (CV, dots) as a function of input variance B_θ, fitted with a Naka-Rushton (NKR) function (dashed curves, parameters shown in light gray). Parameters of the VTF are $\log(n) = 8.4$, $B_{\theta 50} = 14.7°$, $f_0 = 0.4$ for neuron A and $\log(n) = 2.4$, $B_{\theta 50} = 35.0°$, $f_0 = 0.3$ for neuron B. The CV identity curve is shown in solid gray. **c** Histograms of the NKR parameters (in the [5%;95%] range of possible NKR fitting values) for the 249 recorded units. Median values are indicated by a black arrow ($\log(n) = 3.6$, $B_{\theta 50} = 19.2°$, $f_0 = 0.75$).
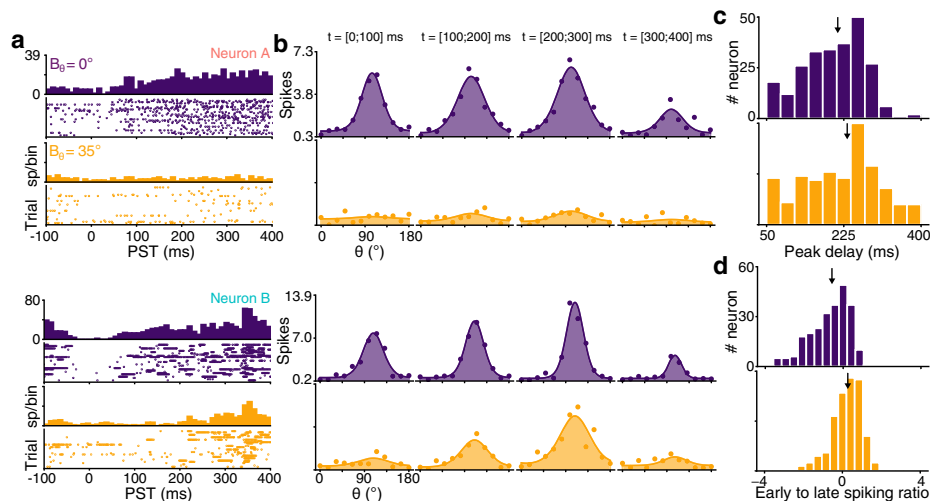
Fig. 3 Neural dynamics depend on input variance. Additional examples are shown in Supplementary Fig. 3. **a** Peristimulus time (PST) histogram and rasterplot for the two previous example neurons, at variance $B_\theta = 0°$ (purple) and $B_\theta = 35°$ (yellow). **b** Tuning curve dynamics in 100 ms windows, starting at labeled times. **c** Delay to peak amplitude of tuning curves for $B_\theta = 0°$ (purple, median = 210 ms) and $B_\theta = 35°$ (yellow, median = 233 ms) for the population. Median values are indicated by a black arrow. **d** Log ratio of early (<100 ms post-stimulation) and late (>200 ms) spike counts for $B_\theta = 0°$ (median = −0.54) and $B_\theta = 35°$ (median = 0.27) for the population.

neurons, including neuron A, and another cluster of 85 neurons associated with neuron B's response type. As neuron B displayed resilience to increased input variance (Fig. 2a), its cluster was labeled resilient neurons. Conversely, neurons clustered with neuron A were labeled vulnerable neurons (blue and red colors, respectively, Fig. 4a). Opting to categorize the data into two distinct response types facilitates a comprehensive understanding of the underlying continuum of behaviors. This approach has proven successful in the characterization of novel visual responses, such as V1 simple/complex cells[24] and MT pattern/component cells[25].

The K-means clustering resulted in a significant difference between the two groups' VTF parameters (Fig. 4b): resilient neurons had significantly more linear modulations ($\log(n)$, $U = 4029.0.0, p < 0.001$, Mann–Whitney U-test), higher changepoints ($B_{\theta 50}, U = 7854.0, p = 0.028$) and better tuning to low-variance inputs ($f_0, U = 4992.0, p < 0.001$), which endows them with the ability to respond to an orientation on a broader range of input variances[26,27]. No significant differences in the variance-HWHH and variance-firing rate functions were observed, except for the non-linearity of the latter metric (Supplementary Fig. 5). This is coherent with the clustering on the statistical measurement of orientation tuning, which showed that resilient neurons remained significantly tuned to higher values of B_θ ($B_{\theta \max}$, Fig. 4c, $U = 9155.0, p < 0.001$). However, both groups of neurons had a similar circular variance for $B_\theta = 35°$ (Fig. 4d). This suggests that both types of neurons were similarly poorly tuned for inputs of the highest variance, but underwent different tuning changes between $B_\theta = 0°$ and $B_\theta = 35°$. In terms of dynamics, the two groups exhibited the same differences that characterized neurons A and B. Resilient neurons discharged significantly later than vulnerable neurons for $B_\theta = 0°$ (Fig. 4e, $U = 8455.5, p = 0.002$), but both groups were on par for inputs of $B_\theta = 35°$ ($U = 7794.5, p = 0.063$). Interestingly, resilient neurons had significantly lower time to the maximum amplitude of the tuning curve for $B\theta = 0°$ (Fig. 4f, $U = 5542.5, p = 0.014$), which opposes the early/late ratio of spikes. Neither group showed

variance-dependent modulation of the delay to maximum spike count ($U = 3058.0, p = 0.084$ and $U = 11545.5, p = 0.090$ for resilient and vulnerable, respectively), and both groups showed similar delay for $B_\theta = 35°$ ($U = 6094.5, p = 0.158$).

The existence of these two groups of neurons could not be attributed to the integration of the drifting motion of the stimuli (direction selectivity index, unused in the clustering process, Fig. 4g, $U = 7031.5, p = 0.910$). Instead, the location of the recorded units (unused in the clustering process) predominantly positioned the resilient neurons in supragranular layers, offering a mechanistic basis for their existence (Fig. 4h). Moreover, resilient neurons have sharper orientation tuning and slower dynamics, which are distinctive features of supragranular neurons[28,29]. This, however, does not establish a functional role for these two types of responses in V1.

Population-level modulations of the orientation code. As the neuronal population has been separated into well-characterized groups, we wish to understand the functional role played by resilient and vulnerable neurons. To that end, we used a neuronal decoder that probes for population codes in V1, enabling us to seek what parameters of the stimuli each neuron group was encoding. We trained a multinomial logistic regression classifier[30], a probabilistic model that classifies data belonging to multiple classes (see "Methods"). This classifier received the firing rate of neurons in a sliding time window (100 ms) and learned, for each neuron, a coefficient that best predicts the class (i.e., the generative parameter θ, B_θ or $\theta \times B_\theta$) of the stimulus.

This decoder was first used to probe for representation of the stimuli's orientations θ in the population activity. For this purpose, the dataset of trials was separated for each variance, such that eight independent, B_θ-specific, orientation decoders were learned, with optimal parametrization (Supplementary Fig. 6). These orientation decoders were able to retrieve the correct stimulus' θ well above the chance level (1 out of 12 orientations, max. accuracy = 10.56 and 4.68 times chance level for $B_\theta = 0°$ and $B_\theta = 35°$, respectively) from the entire population recordings.

ARTICLE
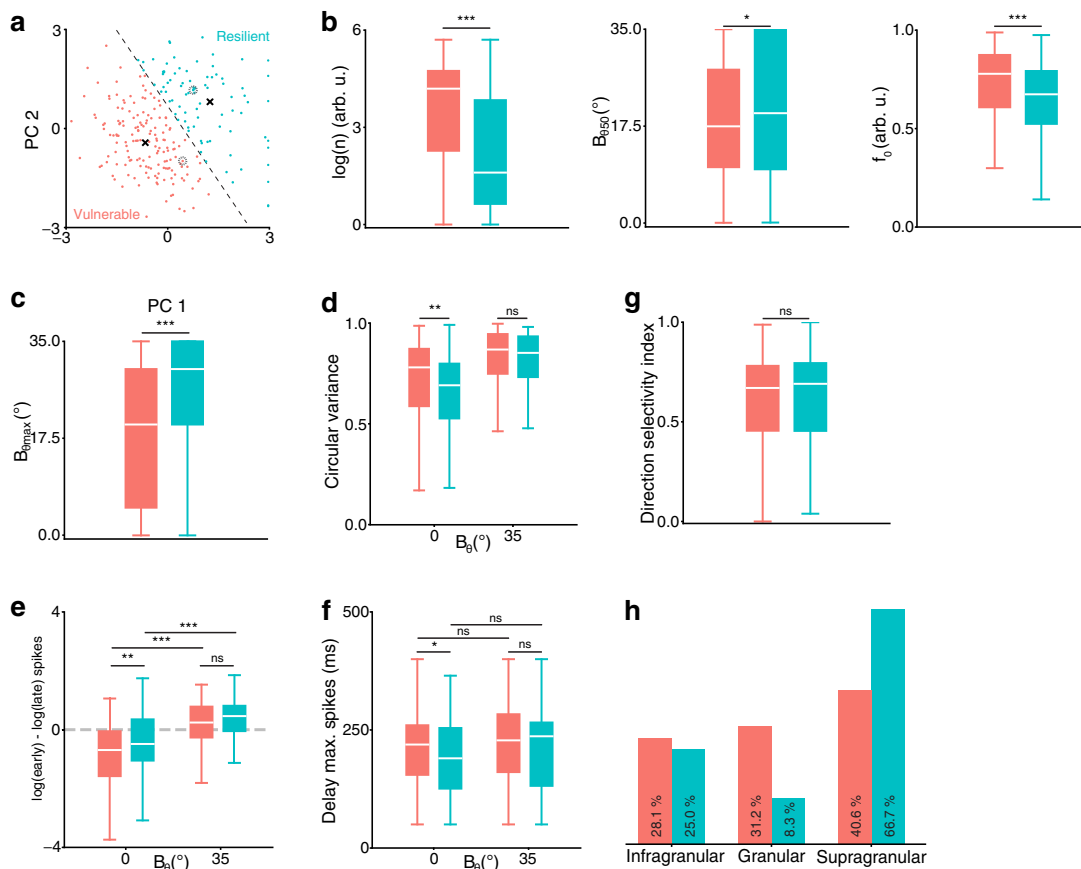
Fig. 4 Responses to changes in variance fall into two categories. a Principal Components (PC) analysis of the data, K-Means clustered (2 clusters, centroids shown as black crosses and separatrix as dashed line). Nine resilient neurons with PC1 > 3 are plotted at PC1 = 3. Neuron A and B are shown as dashed circles. **b** Boxplot of the VTF parameters $\log(n)$, $B_{\theta 50}$, f_0 (ns, not significant; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$ Mann–Whitney U-test). Boxes cover quartile values with a median white line. Whiskers extend to $Q1 - 1.5 * IQR$ and $Q3 + 1.5 * IQR$, where $Q1;Q3$ are lower, upper quartiles and IQR is the interquartile range. **c** Maximum B_θ for significant orientation tuning curve. **d** Circular variance at $B_\theta = 0°$ and $B_\theta = 35°$. **e** Log ratio of the early (<100 ms) and late (>200 ms) spike counts at $B_\theta = 0°$ and $B_\theta = 35°$. **f** Delay to maximum peak amplitude of tuning curves at $B_\theta = 0°$ and $B_\theta = 35°$. **g** Direction selectivity index (unused in the clustering). **h** Laminar position (unused in the clustering).

The temporal evolution of these decoders' accuracy (Fig. 5a) showed that maximally accurate orientation encoding correlates almost linearly with the stimuli's variance, as does the time to reach this accuracy (Fig. 5e, black). These dynamics depend on the input's variance, exhibiting a rapid initial rise followed by a plateau for low-variance inputs, while steadily increasing linearly over time for high-variance inputs. Interestingly, the decoding accuracy remained stable for approximately 100 ms even after a stimulus was no longer displayed. Since the decoders are trained independently in each time window, this accumulative process occurs in the recordings themselves, and not in the decoder.

The full output of these decoders (see "Methods") is a population tuning curve, which displays the likelihood of decoding all possible input classes (here, all θ, Fig. 5b), rather than the proportion of correct decoding reported by the accuracy metric. The clear correlation between the sharpness of these population tuning curves (Fig. 5f left) and the accuracy of the decoder show that improvements in decoding accuracy rely directly on a population-level separation of features within

orientation space[30], particularly at higher B_θ (Fig. 5b, third panel). Overall, B_θ influences the temporality of the orientation code in V1, which echoes its influence on single-neuron dynamics (Fig. 3). The short delay required to process precise inputs is congruent with the feedforward processing latency of V1[31], while the increased time required to reach maximum accuracy for low precision oriented inputs suggests the involvement of a slower, recurrent mechanism.

We then sought to assert the role of the vulnerable and the resilient neural populations by decoding θ from either group. The number of neurons in each group was imbalanced (79 more vulnerable neurons), which influences the accuracy of the decoder (Supplementary Fig. 6). Consequently, we randomly selected (with replacement) groups of 100 neurons from either population, repeating the selection 5 times. Using the same approach as with the global population decoding, we then trained B_θ-specific orientation decoders on the activity of either group of neurons. Resilient neurons outperformed vulnerable ones in decoding accuracy for 56% of the time steps, mainly in the 160–330 ms

103
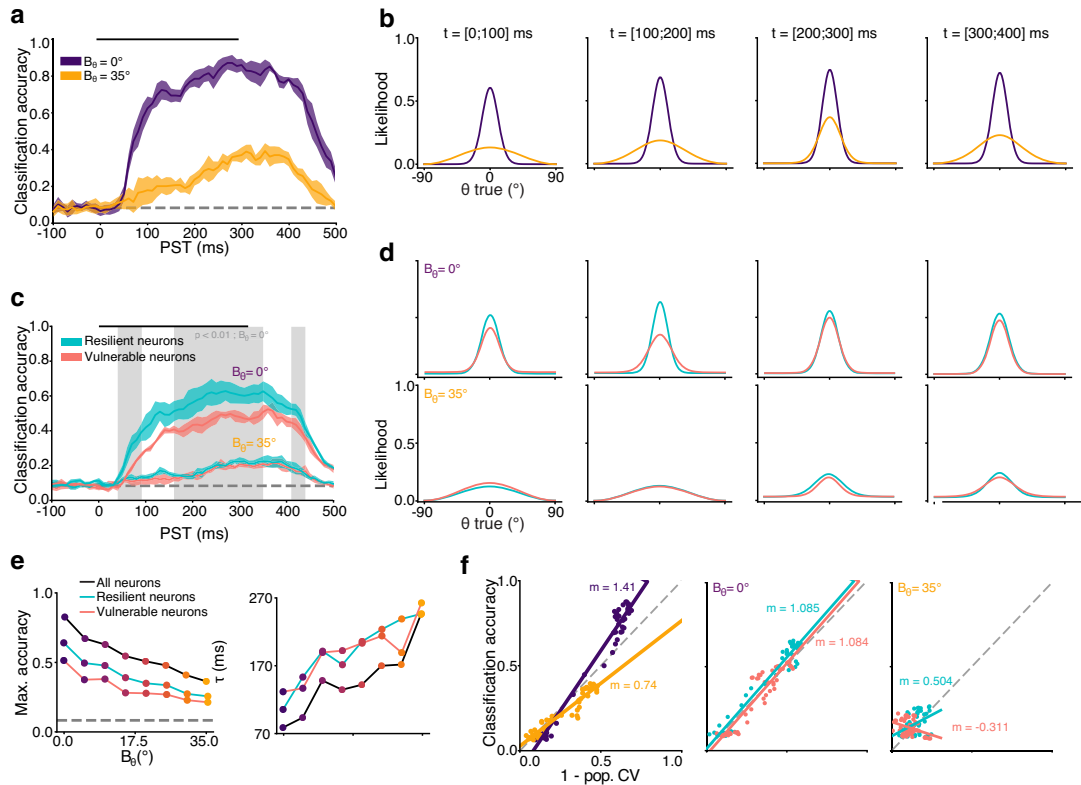
Fig. 5 Input variance modulates orientation decoding in V1. a Time course of orientation θ decoding accuracy at two variances B_θ. Lines are the mean accuracy of $n = 5$ random resampling of 100 neurons and contour the SD. Significantly better decoding from resilient neurons at $B_\theta = 35°$ is shown as a gray overlay (Wilcoxon signed-rank test, threshold $p < 0.01$). Decoding at chance level is represented by a gray dashed line and stimulation time by a black line. **b** Population tuning curves with a von Mises fit, showing the likelihood of decoding each θ in four time windows. **c** Same as (**a**) for the two groups of neurons. **d** Same as (**b**) for the two groups of neurons, with $B_\theta = 0°$ (upper row) and $B_\theta = 35°$ (lower row). **e** Time course parameters for three decoders at all B_θ, estimated by fitting a sigmoid up to PST $= 300$ ms. τ is the time constant. **f** Correlation between classification accuracy and population circular variance for the whole population (left), for both groups with $B_\theta = 0°$ (middle) and $B_\theta = 35°$ (right). Linear regression is shown as solid lines with slope m indicated (all significant, $p < 0.001$, Wald Test with t-distribution).

period (Fig. 5c). However, both groups exhibited similar population tuning curves (Fig. 5d) and time courses (Fig. 5e). Despite the better tuning of resilient neurons to inputs with higher variance (Fig. 4), both groups have overall similar orientation encoding performances for $B_\theta = 35°$. Therefore, orientation can be decoded somewhat more effectively from the resilient neurons at the population level, but neither group appears to have a clear or stable advantage over the other in this regard, especially at higher B_θ.

A subset of V1 neurons co-encode orientation and its variance. Given that orientation encoding did not reveal a fundamental difference in the respective contributions of resilient and vulnerable neurons, we then investigated the encoding of the stimulus' variance B_θ. The same type of decoder previously used failed to infer the variance B_θ (chance level = 1 out of 8 values of B_θ, max. accuracy = 1.91 times chance level) from the population activity (Supplementary Fig. 8a, b). This variance decoding also failed to reach more than twice the chance level (max. accuracy = 1.72 and 1.71 times chance level for resilient and vulnerable

neurons, respectively) in both resilient and vulnerable neurons (Supplementary Fig. 8c,d). At the single neuron level, tuning curves flatten with increments of variance (Supplementary Fig. 2a), which makes it difficult to distinguish activity generated by stimuli with $B_\theta = 0.0°$ and orthogonal orientation from the activity generated by stimuli with $B_\theta = 35.0°$ and preferred orientation. This limitation could potentially stem from the recording scale (249 neurons), which is more than an order of magnitude smaller than the quantity of neurons a single V1 biological decoder can access[32]. Thus, neither the decoding of variance B_θ nor the decoding of orientation θ accounts for a different role between resilient and vulnerable neurons.

The decoding methods used so far have assumed that V1 encodes independently single input parameters. However, a more realistic assumption is to consider the visual system's natural inputs as distributions of information (Fig. 1) that cortical neurons must process from thalamic inputs[33] based on a probabilistic computational principle[34]. Here, this implies that the naturalistic form of processing for a V1 neuron would be co-encoding both the mean feature (θ) and its associated variance (B_θ) to access the entire probability distribution.

ARTICLE
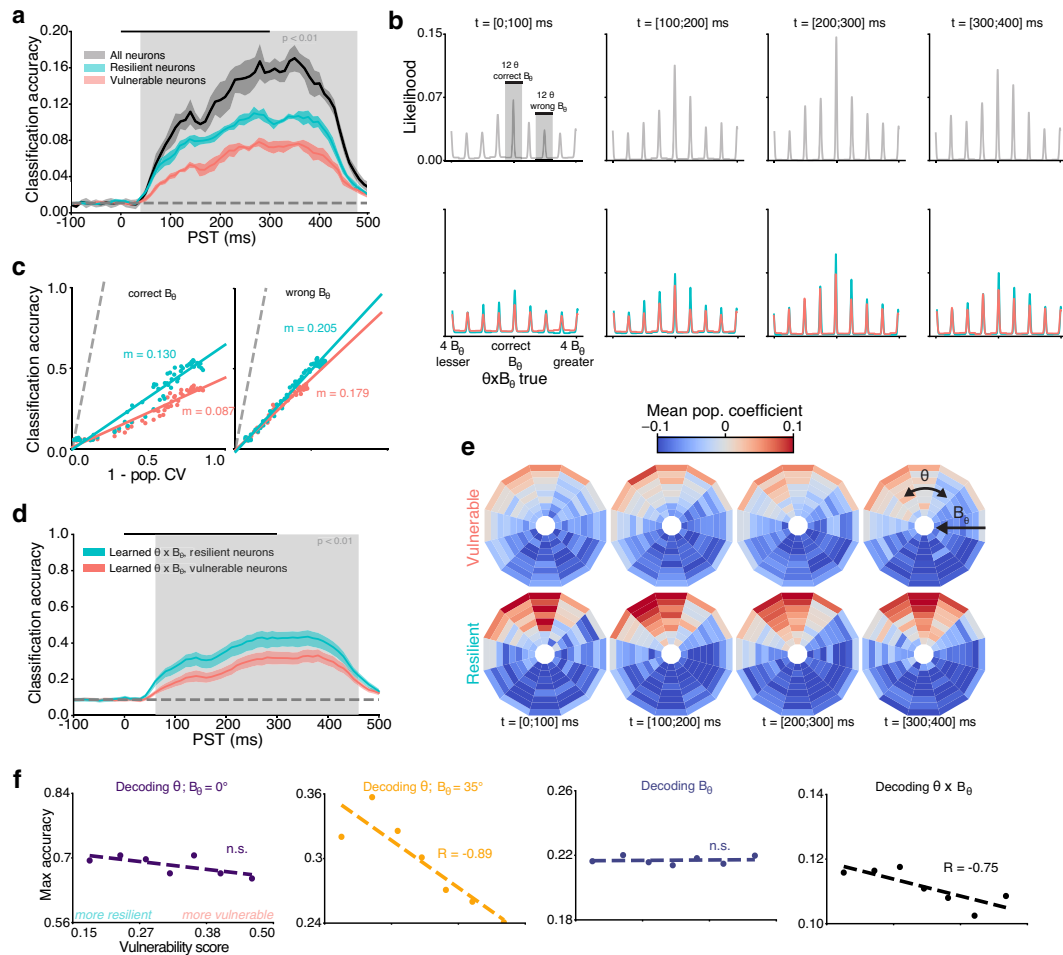
Fig. 6 Orientation and its variance can be decoded from resilient neurons. a Time course of the accuracy for decoding $\theta \times B_\theta$ of Motion Clouds. Lines are the mean accuracy and contour the SD. Significantly better decoding from resilient neurons is shown as a gray overlay (Wilcoxon signed-rank test, threshold $p < 0.01$). Decoding at chance level (1/96) is represented by the gray dashed line. **b** Population tuning curves for the likelihood of decoding each $\theta \times B_\theta$ in four time windows, centered around the correct $\theta \times B_\theta$. **c** Correlation between classification accuracy and population circular variance for correct B_θ population tuning curves (left) and averaged across other B_θ tuning curves (right). Linear regression is shown as solid lines with slope m indicated (all significant $p < 0.001$, Wald Test with t-distribution). **d** Time course of the $\theta \times B_\theta$ decoder, marginalized over B_θ to produce θ-only outputs. **e** Mean decoding coefficients of the two groups yielded from the whole population $\theta \times B_\theta$ decoder. **f** Score-based decoding for θ (first and second columns), B_θ (third) and $\theta \times B_\theta$ (fourth). Raw scores (points) are fitted with a linear regression (dashed curve), with Spearman R shown in the case of a significant correlation ($p < 0.05$).

We thus proceeded to train a decoder that retrieves both orientation and variance of the stimulus' simultaneously, referred to as a $\theta \times B_\theta$ decoder. This decoder correctly predicted orientation and variance with a maximum accuracy reaching 16.36 times the chance level (1/96, Fig. 6a, gray). The likelihood structure (Fig. 6b, upper row) showed that the correct θ was decoded alongside multiple concurrent hypothesis over B_θ. The progressive increase of accuracy stems from the emergence of a dominant encoding of θ at the correct B_θ, consequently diminishing the relative magnitude of representations over other B_θ values over time. Interestingly, resilient neurons showed here a different functional role from vulnerable neurons, with markedly better co-encoding of B_θ and θ (max. accuracy = 11.0 and 9.0

times chance level for resilient and vulnerable neurons, respectively, Fig. 6a, blue, red). Both groups displayed ambiguity regarding B_θ (Fig. 6b, lower row), and correlated sharpening/accuracy ratios on the correct B_θ population curve (Fig. 6c, left) or on the off-median population curves (Fig. 6c, right).

To understand the utility of this co-encoding, we marginalized the decoder over B_θ, creating an orientation-only encoder that simultaneously learned both orientation and variance. Data from resilient neurons then provided significantly better encoding of orientation than vulnerable neurons (max. accuracy = 6.0 and 5.4 times the 1/12 chance level for resilient and vulnerable neurons respectively, Fig. 6d, gray regions), demonstrating that the overall V1 orientation code improves with a co-decoding of its variance.

105

The distinction between resilient and vulnerable neurons is further emphasized by the decoder coefficients, which represent the contributions of each type of neuron toward the overall $\theta \times B_\theta$ code (Fig. 6e, for single neuron examples see Supplementary Fig. 9). Here, these coefficients are depicted as a polar plot, where the orientation θ (centered around preferred orientation) is shown as the angle of each bin from the upper vertical and the variance B_θ is represented as the eccentricity of each bin from the center. Visualizing the coefficients of the whole population decoder (i.e., trained on the 249 neurons, Fig. 6a, gray) shows that the output learned from resilient neurons concurrently informs about both a wide range of orientations and variances, as observed by the extent of the bins in the eccentricity (B_θ) axis (Fig. 6e, bottom row). On the other hand, the decoding process extracted orientation information on a very small range of B_θ from the activity of vulnerable neurons (Fig. 6e, top row). Even though the coefficients are learned independently at each time step, the difference in information between the two groups of neurons remains extremely stable through time.

Overall, orientation and its variance can be co-decoded simultaneously from resilient neurons, while only orientation can be decoded from vulnerable neurons. This is confirmed by a continuous score-based decoding metric based on the K-means parameters (Fig. 6f) that correlates, for the entire population (i.e., without splitting into two groups), their maximum decoding accuracy to a degree of vulnerability/resilience. After providing this functional rationale for resilient and vulnerable neurons, we finally address the question of how both types of neurons can exist in V1.

Recurrent activity can explain the existence of neurons co-encoding orientation and variance. A notable difference between vulnerable and resilient neurons is their different location within the cortical layers (Fig. 4h). This typically implies differences in local circuitry, particularly in the intra-V1 recurrent interactions between cortical columns, which are mostly confined to supragranular layers[35]. Given that resilient neurons are predominantly found in these supragranular layers, we aimed to find a mechanistic rationale for the existence of the two groups of neurons based on local interactions in V1. We developed a neural network from a well-established computational model of recurrent connectivity in V1, originally used to account for the intracortical activity in cat V1[36] and later simplified as a center-surround filter in the orientation domain[29]. This model has already accounted for an extensive range of emerging properties in cortical circuits[37,38]. Briefly, it is built of orientation-selective neurons tiling the orientation space and connected among themselves via recurrent synapses which follow an excitatory/inhibitory difference of von Mises distributions (Fig. 7a). Here, we model inputs with higher variance as more spread in orientation space (Fig. 1) and thus in model space, which hence drives the recurrent dynamics of the model based on B_θ (for a full description, see "Methods").

Considering that feedforward connectivity with heterogeneous tuning can encode mixtures of orientations and natural images[9], we first ran our model without recurrent synapses. We reproduced the heterogeneous selectivity by convolving the input with tuning curves of varying bandwidths (Fig. 7b, inset). This feedforward mode of the network was only able to produce a limited number of responses (Fig. 7b), in which increasing the bandwidth of the tuning curves increased the parameter f_0 of the VTF, but kept n and $B_{\theta 50}$ constant.

Barring that explanation, we focused on the role of recurrent synapses and disabled the convolution of inputs. We varied the concentration parameters of the synaptic distributions κ_{inh} and κ_{exc} (Fig. 7c, e) in 200 even steps ranging from 0.35 to 7, yielding 40,000 possible configurations of the model. This allowed to manipulate the VTF and to accurately reproduce those of single neurons recorded in V1 (neuron A, B in Fig. 2b and C in Supplementary Fig. 1, modeled in Fig. 7c). Altering the type of recurrence between neurons with different orientation preference allowed to reproduce all VTF found in V1. The parameter spaces (Fig. 7e) showed a trend for resilient VTFs (low n, high $B_{\theta 50}$, low f_0) to be found mostly around the K_{exc}; K_{inh} identity line, thus produced by balanced recurrent connectivity. Vulnerable VTFs (high n, low $B_{\theta 50}$, high f_0) were, on the contrary, mostly found above the identity line, where the configuration of the network is dominated by excitation over inhibition. This is consistent with the range of parameters that yielded higher response latency (Fig. 7d), which also occupied more parameter space when input variance increased. In summary, recurrence between V1 neurons seems to be sufficient to explain the existence of vulnerable and resilient neurons and, consequently, to account for the co-encoding of orientation and variance.

Discussion
The variance of oriented inputs to V1 impacts orientation selectivity[9] and we have sought to understand how V1 could process this input parameter. We found that variance causes modulations in tuning (Fig. 2) and dynamics (Fig. 3) of single V1 neurons, which we have classified as either vulnerable or resilient (Fig. 4). Decoding analysis revealed variance-dependent accumulative dynamics in the two groups of neurons (Fig. 5) that are directly tied to a population-level separation of features within orientation space[30]. Both groups can encode orientation but not variance (Supplementary Fig. 8), and only resilient neurons are able to accurately co-encode orientation and variance of the input to V1 (Fig. 6). Based on cortical layer position (Fig. 4h) and on a computational approach (Fig. 7), we propose that the processing input variance in V1 is supported by recurrent connectivity between local cortical populations (Fig. 8). This not only improves the encoding of orientation in V1 but also links directly to canonical Bayesian frameworks, suggesting uncertainty computation as a new mechanism supported by local recurrent cortical connectivity.

Here, we restricted our approach to orientation space, rather than investigating the full extent of spatial relationships which are present in natural images. Thus, full-field stimuli without second-order correlation were used, which compared to a purely ecological environment, have likely excluded end-stopped cells[39]. While this approach limited the responses to V1 and excluded higher-order cortical areas, there exists both neurobiological and computational evidence that V1 does not need to recruit other cortical areas to process orientation variance. For instance, the heterogeneous recurrent excitatory and inhibitory synaptic connectivity in V1[40–43] sustains resilient orientation tuning[44] that can account for the diversity of single neurons' resilience under different connectivity profiles, as explored in our computational model (Fig. 7). This is supported by the temporal scale of local recurrent connectivity, namely the slowly-conducted horizontal waves in an orientation map[45], which fit the view of variance processing as an iterative and accumulative computation implemented by local recurrent interactions between supragranular resilient neurons that are heavily connected through recurrent interactions with neighboring cortical columns[28,29,35,45]. In this regard, our reported time scales may have been slightly affected by the use of anesthesia (halothane), which has a limited visible effect on V1[46,47] and is less likely to cause modulations in this area compared to higher-order areas[48–51].

ARTICLE
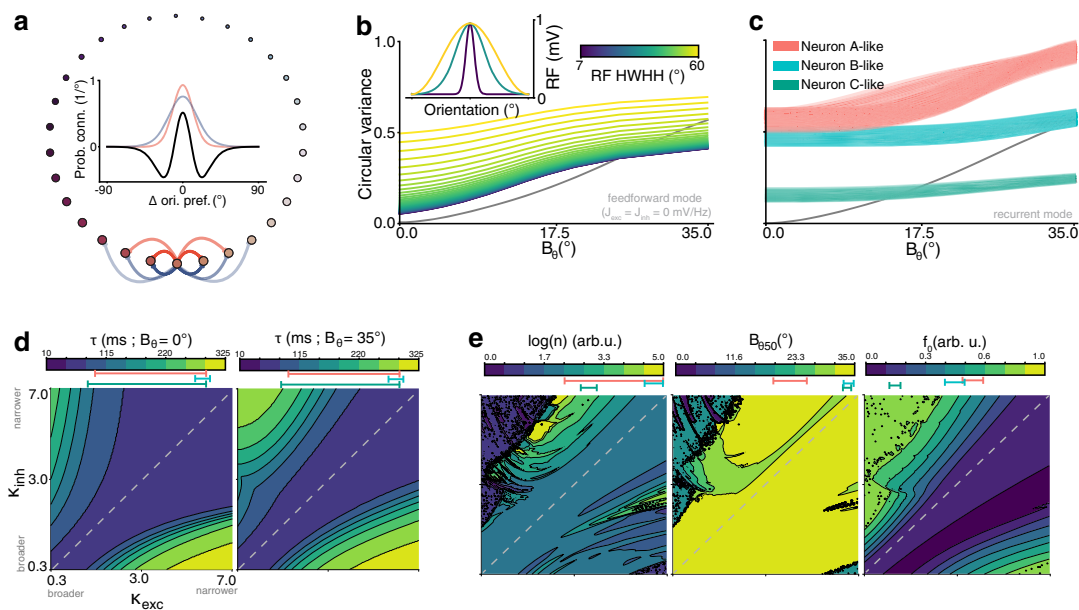
Fig. 7 Recurrent interaction modeling can explain the existence of resilient and vulnerable neurons. a Ring topology of the network, with the preferred orientation of each neuron. Inset: Recurrent connectivity profile for each neuron, computed as a difference (black) of excitatory (red) and inhibitory (blue) profiles, controlled by a measure of concentration κ_{exc} and κ_{inh}, respectively. **b** VTF without recurrent connectivity (i.e., only inputs convolved with receptive fields) and with varying *RF*'s Half-Width at Half-Height (HWHH). The CV identity curve is shown in black. Inset shows examples of receptive fields *RF*. **c** VTF with recurrent connectivity, under two configurations retrieved by searching for VTF parameters close to those of neuron A, B (Fig. 2b) and C (Supplementary Fig. 1). **d** Delay to half maximum firing rate of the model (τ) for each connectivity profile, shown as a contour plot of κ_{exc} and κ_{inh}. [5%;95%] range of the parameters corresponding to the VTFs in (**c**) are displayed below the scale bars. **e** VTF parameters obtained from the model for each connectivity profile, shown as a contour plot of κ_{exc} and κ_{inh}. [5%;95%] range of the parameters of the VTFs shown in (**c**) are displayed below the scale bars.

Computationally, most existing models support the idea that processing orientation variance can be achieved solely with local V1 computations[10]. For instance, Goris et al.[9] reported that heterogeneously tuned V1 populations help encode the orientation distributions found in natural images and that this functional diversity could be accounted for by a linear-nonlinear (L-NL) model. While this could explain the diversity of tuning in our data (Fig. 2), we found that such a model failed to account for some types of modulations of the VTFs (Fig. 7b). Therefore, we employed a model designed to replicate intracortical cat V1 data[38] and demonstrated that it reproduces various VTFs and dynamics observed in our recordings. The model used here pools activity from multiple orientation-tuned units into a single neuron, which we interpreted as a local recurrent model. While our results do not require contributions from extrastriate regions to explain the observed results, the possibility of recurrence involving neurons outside V1 cannot be entirely ruled out at this time[52].

Our study confirms the findings in the anesthetized macaque literature[9] by identifying single-neuron variance modulations that serve as the basis for decoding orientation variance at the population level in V1. This suggests that a common mechanism may underlie this neural mechanism in both felines and primates, which is a fundamental computational requirement for the proper encoding of natural images in V1[53]. Although gain/variance V1 functions have been previously reported[17], we demonstrate a similar input-output relationship in the form of VTFs, that has the added benefit of characterizing and extrapolating

variance modulations across the full dynamical range of V1 populations. Further, we finely analyzed the temporal component of the response that is absent from the literature. We propose that all these response properties can be linked to cortical layers, binding the idea that supragranular neurons with sharp tuning and slow dynamics[28,29] support the co-encoding orientation and its variance.

This leads to an interesting tie to Bayesian inference, namely under the specific case of predictive coding[34], that canonically assigns (inverse) variance weighting of cortical activity to supragranular recurrent connectivity[6,8], without the need for extrastriate computations. This is an interesting perspective that opens up a general interpretation of our results into the broader context of processing variance/precision/uncertainty at different scales of investigations. Extending the present results to other cortical areas or other sensory modalities would be a simple process, given the generative stimulus framework used here[18], which could yield pivotal new insights into our understanding of predictive processes in the brain.

Methods

Visual stimulation. Motion Clouds are generative model-based stimuli[18] that allow for fine parameterized control over naturalistic stimuli[54], which is a desirable trait when probing sensory systems under realistic conditions[21]. They are mathematically defined as band-pass filtered white noise stimuli, whose filters in Fourier space are defined as a parameterized distribution in a given perceptual axis (here, only orientation, but can be extended to speed[55] and scale[56]). Thus, the Motion Clouds presently used are fully characterized by their mean orientation and their

ARTICLE
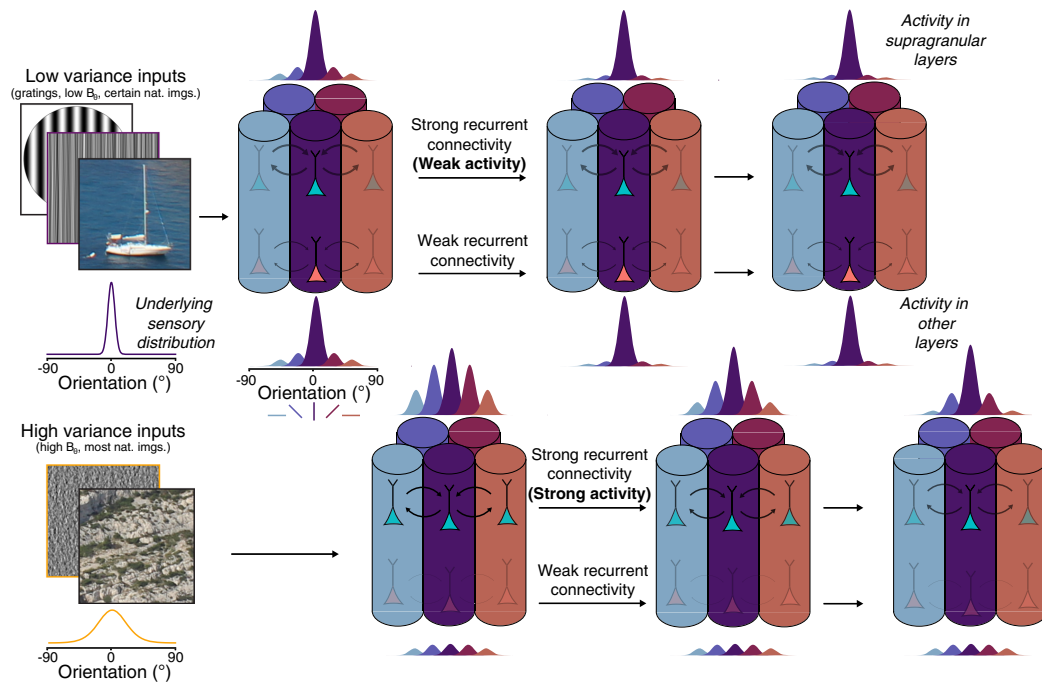
Fig. 8 Summary of the findings. The top row is a representation of a set of orientation-selective units (here, columns) processing low-variance inputs, while the bottom row schematizes the processing of high-variance inputs. In the case of low-variance inputs to V1, the underlying sensory distribution is sharp in the orientation space, driving mostly a single orientation-selective unit that processes orientation in a fast feedforward manner. The feature encoded by this activity then stays stable through time (from left to right). For inputs of higher orientation variance, the sensory input is broadly distributed in the orientation space, which drives many dissimilarly tuned units, thus recruiting slow recurrent interactions. The quality of feature encoding progressively increases through time, as recurrent interactions perform computations to represent the most salient oriented feature in the input.

orientation variance, such that a given stimulus S can be defined as:

$$S = \mathcal{F}^{-1}(O(\theta, B_\theta)) \quad (1)$$

where $\mathcal{F}$ is the Fourier transform and O the orientation envelope, characterized by its mean orientation θ and its orientation bandwidth B_θ. For $B_\theta < 45.0°$, $B_\theta = 1/\sqrt{\kappa}$, where κ is the concentration parameter of a von Mises distribution, and hence approximates the standard deviation[57]. It thus serves as a measure of the orientation variability in the pattern, and as such, we used the term variance to describe it throughout the text. A total of 96 different stimuli were generated, with 12 mean orientations θ ranging from 0 to π in even steps, and eight orientation variance B_θ ranging from ≈0 to $\pi/5$ in even steps. The orientation envelope is a von Mises distribution:

$$O(\theta, B_\theta) = \exp\left\{\frac{\cos(2(\theta_f - \theta))}{4 \cdot B_\theta^2}\right\} \quad (2)$$

where θ_f is the angle of the frequency components of the envelope in the Fourier plane, which controls the spatial frequency parameters of the stimuli, set here at 0.9 cycle per degree. The stimuli were drifting orthogonally in either direction with respect to the mean orientation θ at a speed of 10°/s, which is optimal to drive V1 neurons[58]. For the range of values of B_θ considered here, the orientation envelope approximates a Gaussian distribution and B_θ is thus a measure of the variance of the orientation content of the stimuli.

All stimuli were generated using open-source Python code (see Additional information) and displayed using Psychopy[59]. Monocular stimuli were projected with a ProPixx projector (VPixx Technologies Inc.) onto an isoluminant screen (Da-Lite©) covering 104° × 79° of visual angle. All stimuli were displayed for 300 ms, interleaved with a mean luminance screen (25 cd/m²) shown for 150 ms between each trial. Trials were fully randomized, and each stimulus (a unique combination of $\theta \times B_\theta \times$ drift direction) was presented 15 times. Stimuli were shown at 100% contrast, meaning that as B_θ increased, the amount of orientation energy at median orientation θ decreased, and conversely for off-median orientations (as illustrated in Fig. 1b). This differs from manipulating the contrast, which would reduce the orientation energy at all orientations.

Surgery. Experiments were conducted on three adult cats (3.6–6.0 kg, 2 males). All surgical and experimental procedures were carried out in compliance with the guidelines of the Canadian Council on Animal Care and were approved by the Ethics Committee of the University of Montreal (CDEA #20-006). Animals were initially sedated using acepromazine (Atravet®, 1 mg/kg) supplemented by atropine (0.1 mg/kg). Anesthesia was induced with 3.5% isoflurane in a 50:50 mixture of $O_2:N_2O$ (v/v). Following tracheotomy, animals underwent artificial ventilation as muscle relaxation was achieved and maintained with an intravenous injection of 2% gallamine triethiodide (10 mg/kg/h) diluted in a 1:1 (v/v) solution of 5% dextrose lactated Ringer solution. Through the experiment, the expired level of CO_2 was maintained between 35 and 40 mmHg by adjusting the tidal volume and respiratory rate. Heart rate was monitored and body temperature was maintained at 37 °C by means of a feedback-controlled heated blanket. Lidocaine hydro-chlorine (2%) was applied locally at all incisions and pressure points and a cra-niotomy was performed over area 17 (V1, Horsley-Clarke coordinates 4-8P; 0.5–2 L). Dexamethasone (4 mg) was administered intramuscularly every 12 h to reduce cortical swelling. Eye lubricant was regularly applied to avoid corneal dehydration.

Electrophysiological recordings. During each recording session, pupils were dilated using atropine (Mydriacyl) while nictitating membranes were retracted using phenylephrine (Mydfrin). Rigid contact lenses of appropriate power were used to correct the eyes' refraction. Anesthesia was changed to 0.5–1% halothane to avoid anesthesia-induced modulation of visual responses[47]. Finally, small dur-ectomies were performed before each electrode insertion and a 2% agar solution in saline was applied over the exposed cortical surface to stabilize recordings. Linear probes (≈1 MΩ, 1x32-6mm-100-177, Neuronexus) were lowered in the cortical tissue perpendicularly to the pia, and extracellular activity was acquired at 30KHz using an Open Ephys acquisition board[60]. Single units were isolated using Kilosort 2[61] and manually curated using Phy[62]. Clusters with low amplitude templates or ill-defined margins were excluded from further analysis. The additional exclusion was performed if a cluster was unstable (firing rate below 5 spikes.s⁻¹ for more than 30 s), or if the neuron was not deemed sufficiently orientation selective ($R^2 < 0.75$ when fitted with a von Mises distribution). Passing that exclusion step,

108

ARTICLE

all remaining neurons responded to Motion Clouds. Laminar positions were determined by the depth of the recording site with respect to the pia, which was then cross-validated by the evoked Local Field Potential (LFP) using sink/source analysis[63,64].

Single neuron analysis. Orientation tuning curves were computed by selecting a 300 ms window maximizing spike-count variance[65]. The firing rate was averaged across drift directions and a von Mises distribution[57] was fitted to the data:

$$f(\theta_k) = R_0 + (R_{max} - R_0) \cdot \exp\left\{\kappa \cdot (\cos(2(\theta_k - \theta_{pref})) - 1)\right\} \quad (3)$$

where θ_k is the orientation of the stimuli, R_{max} is the response (baseline subtracted) at the preferred orientation θ_{pref}, R_0 the response at the orientation orthogonal to θ_{pref} and κ a measure of concentration. To control for direction selectivity when averaging tuning curves across drift direction, we computed a direction selectivity index:

$$D_s = \frac{R_{pref} - R_{null}}{R_{pref}} \quad (4)$$

where R_{pref} is the firing rate at the preferred direction (baseline subtracted) and R_{null} is the firing rate at the preferred direction plus π. The quality of each tuning curve was assessed by computing a global metric, the circular variance (CV) of the unfitted data, which varies from 0 for perfectly orientation-selective neurons to 1 for orientation-untuned neurons[29]. It is defined as:

$$CV = 1 - \left| \frac{\sum_k R(\theta_k) \cdot \exp\{2i\theta_k\}}{\sum_k R(\theta_k)} \right| \quad (5)$$

where $R(\theta_k)$ is the response of a neuron (baseline subtracted) to a stimulus of angle θ_k. The changes of CV as a function of B_θ were fitted with a Naka-Rushton function[22]:

$$f(B_\theta) = f_0 + f_{max} \frac{B_\theta^n}{B_\theta^n + B_{\theta 50}^n} \quad (6)$$

where f_0 is the base value of the function, $f_0 + f_{max}$ its maximal value, $B_{\theta 50}$ the stimulus' variance at half f_{max} and n a strictly positive exponent of the function.

The significance of the tuning to orientation was measured by comparing the unfitted firing rate at the preferred and orthogonal orientations across trials, using a Wilcoxon signed-rank test correct for continuity, and the maximum value of B_θ which yielded a significant result was designed as $B_{\theta max}$ (i.e., the maximum variance at which a neuron is still tuned). Shifts of the preferred orientation were evaluated as the difference of θ_{pref} between trials where $B_\theta = 0°$ and $B_\theta = B_{\theta max}$. The significance of the variation of the peak amplitude of the tuning curve was measured by comparing the unfitted firing rate at the preferred orientation between trials where $B_\theta = 0°$ and $B_\theta = B_{\theta max}$.

Population decoding. The parameters used to generate Motion Clouds were decoded from the neural recordings using a multinomial logistic regression classifier[30]. For a given stimulus, the activity of all the recorded neurons was a vector $X(t) = \begin{bmatrix} X_1(t) & X_2(t) & \cdots & X_{249}(t) \end{bmatrix}$, where $X_i(t)$ is the spike count of neuron i in a time window $[t; t + \Delta T]$. The onset of this window t was slid from -200 to 400 ms (relative to the stimulation time) in steps of 10 ms while ΔT was kept constant at 100 ms. It should be noted that merging neural activity across electrodes or experiments is a common procedure[66,67], which we validated in our data by verifying that the electrode or experiment which yielded the data could not be decoded from the neural activity (Supplementary Fig. 7). Mathematically, the multinomial logistic regression is an extension of the binary logistic regression[30] trained here to classify the spike vector $X(t)$ between K classes. The probability of any such vector belonging to a given class is:

$$P(y = k | X(t)) = \frac{\exp\left\{\langle \beta_k, X(t) \rangle\right\}}{\sum_{k'=1}^{K} \exp\left\{\langle \beta_{k'}, X(t) \rangle\right\}} \quad (7)$$

where $\langle \cdot, \cdot \rangle$ is the scalar product over the different neurons, $k = 1, …, K$ is the class out of K possible values and β_k are the coefficients learned during the training procedure of the classifier. Several decoders were trained with classification tasks: decoding orientation θ ($K = 12$, Fig. 5), decoding orientation variance B_θ ($K = 8$, Supplementary Fig. 8) or both ($K = 12 \times 8 = 96$, Fig. 6). All meta-parameters were controlled, showing that the decoding performances stem mainly from experimental data rather than fine-tuning of the decoder parameterization (Supplementary Fig. 6). For all decoding experiments reported, we used integration window size $\Delta T = 100$ ms, penalty type $= \ell_2$, regularization strength $C = 1$. and train/test split size $= 0.15$.

The performance of all decoders was reported as the average accuracy across all classes K, known as the balanced accuracy score[68]. The accuracy for each specific class k can also be reported in the form of a population tuning curve, in which the likelihood of decoding each possible class K is given by equation (7). The significance of differences between two neuron groups was reported only when two consecutive time steps, i.e., 20 ms or more, exhibited significant differences. To estimate the time course of the decoders, they were fitted in the [0; 300] ms range

with a sigmoid function:

$$\sigma = max_{acc}\left(\frac{1}{1 + e^{-k\tau}}\right) + min_{acc} \quad (8)$$

where max_{acc} and min_{acc} are respectively the maximum and minimum accuracies of the decoder, k the steepness and τ the time constant of the function. To perform decoding on the same number of vulnerable or resilient neurons, we randomly picked replacement groups of 100 neurons and bootstrapped this process five times.

As the neurons were clustered into two populations for comparison purposes (Fig. 4), we also reported the decoding accuracy based on a continuous vulnerability score (Fig. 6f). This score was computed as a sum of neuronal responses variables significantly different after the clustering, weighted by their mean Principal Component (PC-1 and PC-2) parameters:

$$\begin{aligned} score = 1 &- W_1(B_{\theta 50}) + W_2(1 - \log(n)) + W_3(1 - f_0) + W_4(B_{\theta max}) \\ &+ W_5(1 - CV) + W_6(\text{early/late ratio}) + W_7(\text{delay}) \end{aligned} \quad (9)$$

where W_i is a parameter yielded by the Principal Component Analysis corresponding to its associated neuronal response variable. Each variable is normalized, yielding a scalar score that varies between 0 (most resilient) to 1 (most vulnerable neuron). This score-based decoding was performed on groups of 100 neurons sorted by descending score and repeated a total of seven times on increasingly more vulnerable neurons (thus with an overlap of 20 neurons).

Computational model. We used a recurrent network of orientation-tuned neurons to model responses to increasing orientation variance B_θ. The model presently used was first used to account for the intracortical activity in the cat primary visual cortex[36], although it was presently simplified as a center-surround filter in the orientation domain[29]. Notably, this network has been able to account for numerous experimental findings, including learning and adaptation of cortical neurons[37,38], whose implementations are similar to ours.

The model consisted of N orientation-tuned neurons, evenly tiling the orientation space between $-\pi$ and π. Each neuron is modeled as a single passive unit whose membrane potential obeys the equation:

$$\tau \delta V / \delta T + V = V_{ff} + V_{exc} - V_{inh} \quad (10)$$

where τ is the membrane time constant and V_{ff}, V_{exc}, V_{inh} are the synaptic potentials coming from the feedforward input, recurrent excitatory and recurrent inhibitory connectivity, respectively. The firing rate R at time t of each neuron is computed as an instantaneous quantity modulated by a gain α:

$$R(t) = \alpha \cdot max(V(t), 0) \quad (11)$$

For computational simplicity, the neurons had no spontaneous firing rate and V was measured relative to the firing threshold. Each neuron could send mixed excitatory and inhibitory synaptic potentials to its neighbor, although this specific model has been reported to achieve similar behavior with separate units[38]. For each stimulus of main orientation θ, the input to a cell with preferred orientation θ_{pref} is:

$$V_{ff}(\theta_{pref}) = J_{ff} \frac{e^{\kappa_{ff} \cdot \cos(2(\theta - \theta_{pref}))}}{2\pi I_0(\kappa_{ff})} \quad (12)$$

where J_{ff} is the strength of the input and I_0 is the modified Bessel function of order 0. The right-hand side of the equation describes a von Mises with mean θ_{pref} and concentration κ_{ff}. This latter parameter is related to the orientation variance B_θ, which was varied to yield a model's TVF B_θ/CV curves:

$$B_\theta = \sqrt{\frac{0.5 \arccos((\log(0.5) + \kappa_{ff})/\kappa_{ff})}{2 \log(2)}} \quad (13)$$

a total of 20 B_θ spanning the same range used in the experiments were used, each with 32 different θ tiling a $[-75°; 75°]$ orientation space. The recurrent connectivity profile for excitatory (C_{exc}) and inhibitory (C_{inh}) synapses was controlled by separate von Mises distributions over the orientation space Θ:

$$C_{exc}(\theta_{pref}) = \frac{e^{\kappa_{exc} \cdot \cos(2(\Theta - \theta_{pref}))}}{2\pi I_0(\kappa_{exc})} \quad (14)$$

$$C_{inh}(\theta_{pref}) = \frac{e^{\kappa_{inh} \cdot \cos(2(\Theta - \theta_{pref}))}}{2\pi I_0(\kappa_{inh})} \quad (15)$$

which are both used to describe an overall connectivity kernel:

$$C_{tot}(\theta_{pref}) = J_{exc}C_{exc} - J_{inh}C_{inh} \quad (16)$$

which followed a typical Ricker wavelet (or Mexican hat) shape (Fig. 7d). The overall activity of the network is then a weighted sum of the firing rates of all the neurons:

$$V_{exc} - V_{inh}(t) = \sum_{\Theta} C_{tot}(\theta_{pref}) \cdot R(t) \quad (17)$$

Parameterization of the model was done to match single V1 neuron recordings of anesthetized cats, in an experimental setup similar to the one used here[69].

ARTICLE

The computational procedure to match experimental data was entirely done in a previous publication[38]. Briefly, it consisted in scanning a range of possible values for each parameter, then finding all possible combinations using a metric of likeliness to single grating response, time-to-peak, peak response and tuning width. The parameters yielded by this procedure were $\tau = 10.8$ ms; $\alpha = 10.6$ Hz/mV; $J_{\text{ff}} = 9.57$ mv/Hz; $J_{\text{exc}} = 1.71$ Hz/mV; $J_{\text{inh}} = 2.0178$ Hz/mV. For the feedforward mode of the model (Fig. 7b), J_{exc} and J_{inh} were set to 0 Hz/mV and the input was convolved with a receptive field:

$$\text{RF} = \frac{e^{\kappa_{\text{RF}} \cdot \cos(2(\theta - \Theta_{\text{pref}}))}}{2\pi I_0(\kappa_{\text{RF}})} \qquad (18)$$

of which we reported the Half-Width at Half-Height, given by[70]:

$$\text{HWHH} = 0.5 \arccos(\frac{\log(0.5) + \kappa}{\kappa}) \qquad (19)$$

For the recurrent mode (Fig. 7c–e), the concentration measures of the recurrent connectivity profiles κ_{exc} and κ_{inh} were both varied from 0.35 to 7, in 200 even steps, and the input was not convolved with a receptive field.

Statistics and reproducibility. All data were analyzed using custom Python code. Statistical analysis was performed using non-parametric tests. Wilcoxon signed-rank test with discarding of zero-differences was used for paired samples and Mann–Whitney U-test with exact computation of the U distribution was used for independent samples. Due to the impracticality of using error bars when plotting time series, colored contours are used to represent standard deviation values (unless specified otherwise), with a solid line representing mean values. For box-plots, the box extends from the lower to upper quartile values, with a solid white line at the median value. The upper and lower whiskers extend to respectively $Q1 - 1.5*IQR$ and $Q3 + 1.5*IQR$, where $Q1$ and $Q3$ are the lower and upper quartiles and IQR is the interquartile range.

Reporting summary. Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data used in the present study is publicly available in a Figshare repository[71]. Unprocessed electrophysiological recording files are available upon reasonable request to the corresponding author.

Code availability

Custom Python code written for the present study is publicly available in a GitHub repository[72].

References

1. Hubel, D. H. & Wiesel, T. N. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* **148**, 574 (1959).
2. Priebe, N. J. Mechanisms of orientation selectivity in the primary visual cortex. *Annu. Rev. Vis. Sci.* **2**, 85–107 (2016).
3. Fiser, J., Chiu, C. & Weliky, M. Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature* **431**, 573–578 (2004).
4. Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001).
5. Helmholtz, H. v. *Helmholtz's Treatise on Physiological Optics,* 3 Vols. (Optical Society of America, 1924).
6. Friston, K. A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* **360**, 815–836 (2005).
7. Barthelmé, S. & Mamassian, P. Evaluation of objective uncertainty in the visual system. *PLoS Comput. Biol.* **5**, e1000504 (2009).
8. Bastos, A. M. et al. Canonical microcircuits for predictive coding. *Neuron* **76**, 695–711 (2012).
9. Goris, R. L., Simoncelli, E. P. & Movshon, J. A. Origin and function of tuning diversity in macaque visual cortex. *Neuron* **88**, 819–831 (2015).
10. Orbán, G., Berkes, P., Fiser, J. & Lengyel, M. Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron* **92**, 530–543 (2016).
11. Festa, D., Aschner, A., Davila, A., Kohn, A. & Coen-Cagli, R. Neuronal variability reflects probabilistic inference tuned to natural image statistics. *Nat. Commun.* **12**, 3635 (2021).
12. Keeble, D., Kingdom, F., Moulden, B. & Morgan, M. Detection of orientationally multimodal textures. *Vision Res.* **35**, 1991–2005 (1995).
13. Beaudot, W. H. & Mullen, K. T. Orientation discrimination in human vision: psychophysics and modeling. *Vision Res.* **46**, 26–46 (2006).
14. Phillips, G. C. & Wilson, H. R. Orientation bandwidths of spatial mechanisms measured by masking. *J. Opt. Soc. Am. A* **1**, 226–232 (1984).
15. Heeley, D., Timney, B., Paterson, I. & Thompson, R. Width discrimination for band-pass stimuli. *Vision Res.* **29**, 901–905 (1989).
16. Heeley, D. W. & Buchanan-Smith, H. M. The influence of stimulus shape on orientation acuity. *Exp. Brain Res.* **120**, 217–222 (1998).
17. Hénaff, O. J., Boundy-Singer, Z. M., Meding, K., Ziemba, C. M. & Goris, R. L. Representation of visual uncertainty through neural gain variability. *Nat. Commun.* **11**, 2513 (2020).
18. Leon, P. S., Vanzetta, I., Masson, G. S. & Perrinet, L. U. Motion clouds: model-based stimulus synthesis of natural-like random textures for the study of motion perception. *J. Neurophysiol.* **107**, 3217–3226 (2012).
19. Johnson, A. P. & Baker, C. L. First-and second-order information in natural images: a filter-based approach to image statistics. *J. Opt. Soc. Am. A* **21**, 913–925 (2004).
20. Field, D. J. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* **4**, 2379–2394 (1987).
21. Rust, N. C. & Movshon, J. A. In praise of artifice. *Nat. Neurosci.* **8**, 1647–1650 (2005).
22. Naka, K. & Rushton, W. A. S-potentials from colour units in the retina of fish (Cyprinidae). *J. Physiol.* **185**, 536–555 (1966).
23. Thorndike, R. L. Who belongs in the family? *Psychometrika* **18**, 267–276 (1953).
24. Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106 (1962).
25. Movshon, J. A. The analysis of moving visual patterns. *Exp. Brain Res.* **54**, 117–151 (1985).
26. Laughlin, S. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforschung C Biosci.* **36**, 910–912 (1981).
27. Kinouchi, O. & Copelli, M. Optimal dynamical range of excitable networks at criticality. *Nat. Phys.* **2**, 348–351 (2006).
28. Ringach, D. L., Hawken, M. J. & Shapley, R. Dynamics of orientation tuning in macaque primary visual cortex. *Nature* **387**, 281–284 (1997).
29. Ringach, D. L., Shapley, R. M. & Hawken, M. J. Orientation selectivity in macaque V1: diversity and laminar dependence. *J. Neurosci.* **22**, 5639–5651 (2002).
30. Bishop, C. M. *Pattern Recognition and Machine Learning* (Springer, 2006).
31. Berens, P. et al. A fast and simple population code for orientation in primate V1. *J. Neurosci.* **32**, 10618–10626 (2012).
32. Chavane, F., Perrinet, L. U. & Rankin, J. Revisiting horizontal connectivity rules in V1: from like-to-like towards like-to-all. *Brain Struct. Funct.* **227**, 1279–1295 (2022).
33. Roelfsema, P. R., Engel, A. K., König, P. & Singer, W. Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature* **385**, 157–161 (1997).
34. Aitchison, L. & Lengyel, M. With or without you: predictive coding and Bayesian inference in the brain. *Curr. Opin. Neurobiol.* **46**, 219–227 (2017).
35. Douglas, R. J., Martin, K. A. & Whitteridge, D. A canonical microcircuit for neocortex. *Neural Comput.* **1**, 480–488 (1989).
36. Somers, D. C., Nelson, S. B. & Sur, M. An emergent model of orientation selectivity in cat visual cortical simple cells. *J. Neurosci.* **15**, 5448–5465 (1995).
37. Teich, A. F. & Qian, N. Learning and adaptation in a recurrent model of V1 orientation selectivity. *J. Neurophysiol.* **89**, 2086–2100 (2003).
38. del Mar Quiroga, M., Morris, A. P. & Krekelberg, B. Adaptation without plasticity. *Cell Rep.* **17**, 58–68 (2016).
39. Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J. Neurophysiol.* **28**, 229–289 (1965).
40. Jia, H., Rochefort, N. L., Chen, X. & Konnerth, A. Dendritic organization of sensory input to cortical neurons in vivo. *Nature* **464**, 1307–1312 (2010).
41. Chen, X., Leischner, U., Rochefort, N. L., Nelken, I. & Konnerth, A. Functional mapping of single spines in cortical neurons in vivo. *Nature* **475**, 501–505 (2011).
42. Iacaruso, M. F., Gasler, I. T. & Hofer, S. B. Synaptic organization of visual space in primary visual cortex. *Nature* **547**, 449–452 (2017).
43. Scholl, B., Wilson, D. E. & Fitzpatrick, D. Local order within global disorder: synaptic architecture of visual space. *Neuron* **96**, 1127–1138 (2017).
44. Monier, C., Chavane, F., Baudot, P., Graham, L. J. & Frégnac, Y. Orientation and direction selectivity of synaptic inputs in visual cortical neurons: a diversity of combinations produces spike tuning. *Neuron* **37**, 663–680 (2003).
45. Chavane, F. et al. Lateral spread of orientation selectivity in V1 is controlled by intracortical cooperativity. *Front. Syst. Neurosci.* **5**, 4 (2011).
46. Uhl, R. R., Squires, K. C., Bruce, D. L. & Starr, A. Effect of halothane anesthesia on the human cortical visual evoked response. *Anesthesiology* **53**, 273–276 (1980).

ARTICLE

47. Villeneuve, M. Y. & Casanova, C. On the use of isoflurane versus halothane in the study of visual response properties of single cells in the primary visual cortex. *J. Neurosci. Methods* **129**, 19–31 (2003).

48. Martinez-Conde, S. et al. Effects of feedback projections from area 18 layers 2/3 to area 17 layers 2/3 in the cat visual cortex. *J. Neurophysiol.* **82**, 2667–2675 (1999).

49. Wang, C., Waleszczyk, W., Burke, W. & Dreher, B. Modulatory influence of feedback projections from area 21a on neuronal activities in striate cortex of the cat. *Cereb. Cortex* **10**, 1217–1232 (2000).

50. Huang, L., Chen, X. & Shou, T. Spatial frequency-dependent feedback of visual cortical area 21a modulating functional orientation column maps in areas 17 and 18 of the cat. *Brain Res.* **998**, 194–201 (2004).

51. Hudetz, A. G., Vizuete, J. A., Pillay, S. & Mashour, G. A. Repertoire of mesoscopic cortical activity is not reduced during anesthesia. *Neuroscience* **339**, 402–417 (2016).

52. Carandini, M. et al. Do we know what the early visual system does? *J. Neurosci.* **25**, 10577–10597 (2005).

53. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).

54. Vacher, J., Meso, A. I., Perrinet, L. U. & Peyré, G. Biologically inspired dynamic textures for probing motion perception. In *Proc. Twenty-Ninth Annual Conference on Neural Information Processing Systems (NIPS)* (NIPS, 2015).

55. Simoncini, C., Perrinet, L. U., Montagnini, A., Mamassian, P. & Masson, G. S. More is not always better: adaptive gain control explains dissociation between perception and action. *Nat. Neurosci.* **15**, 1596–1603 (2012).

56. Ravello, C. R., Perrinet, L. U., Escobar, M.-J. & Palacios, A. G. Speed-selectivity in retinal ganglion cells is sharpened by broad spatial frequency, naturalistic stimuli. *Sci. Rep.* **9**, 1–16 (2019).

57. Swindale, N. V. Orientation tuning curves: empirical description and estimation of parameters. *Biol. Cybern.* **78**, 45–56 (1998).

58. Movshon, J. A., Thompson, I. & Tolhurst, D. Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *J. Physiol.* **283**, 101–120 (1978).

59. Peirce, J. et al. Psychopy2: experiments in behavior made easy. *Behav. Res. Methods* **51**, 195–203 (2019).

60. Siegle, J. H. et al. Open Ephys: an open-source, plugin-based platform for multichannel electrophysiology. *J. Neural Eng.* **14**, 045003 (2017).

61. Pachitariu, M., Steinmetz, N. A., Kadir, S. N., Carandini, M. & Harris, K. D. Fast and accurate spike sorting of high-channel count probes with kilosort. *Adv. Neural Inf. Process. Syst.* **29**, 4448–4456 (2016).

62. Rossant, C. et al. Spike sorting for large, dense electrode arrays. *Nat. Neurosci.* **19**, 634–641 (2016).

63. Katzner, S. et al. Local origin of field potentials in visual cortex. *Neuron* **61**, 35–41 (2009).

64. Maier, A., Adams, G. K., Aura, C. & Leopold, D. A. Distinct superficial and deep laminar domains of activity in the visual cortex during rest and stimulation. *Front. Syst. Neurosci.* **4**, 31 (2010).

65. Smith, M. A., Majaj, N. J. & Movshon, J. A. Dynamics of motion signaling by neurons in macaque area MT. *Nat. Neurosci.* **8**, 220–228 (2005).

66. Quiroga, R. Q., Reddy, L., Koch, C. & Fried, I. Decoding visual inputs from multiple neurons in the human temporal lobe. *J. Neurophysiol.* **98**, 1997–2007 (2007).

67. Guitchounts, G., Masis, J., Wolff, S. B. & Cox, D. Encoding of 3D head orienting movements in the primary visual cortex. *Neuron* **108**, 512–525 (2020).

68. Brodersen, K. H., Ong, C. S., Stephan, K. E. & Buhmann, J. M. The balanced accuracy and its posterior distribution. In *Proc. 2010 20th International Conference on Pattern Recognition*, 3121–3124 (IEEE, 2010).

69. Felsen, G. et al. Dynamic modification of cortical orientation tuning mediated by recurrent connections. *Neuron* **36**, 945–954 (2002).

70. Swindale, N. V., Grinvald, A. & Shmuel, A. The spatial pattern of response magnitude and selectivity for orientation and direction in cat visual cortex. *Cereb. Cortex* **13**, 225–238 (2003).

71. Ladret, H. Data for Ladret et al. 2023 : Cortical recurrence supports resilience to sensory variance in the primary visual cortex. figshare https://figshare.com/articles/dataset/Data_for_Ladret_et_al_2023_Cortical_recurrence_supports_resilience_to_sensory_variance_in_the_primary_visual_cortex_/23366588 (2023).

72. Ladret, H. hugoladret/variance-processing-V1: v1.0-publication. Zenodo https://doi.org/10.5281/zenodo.8016705 (2023).

Author contributions
L.U.P., C.C., F.C., N.C., and H.J.L. designed the study. H.J.L, N.C. and L.I. collected the data. H.J.L. and L.U.P. analyzed the data. H.J.L. and L.U.P. wrote the original draft of the manuscript. All authors reviewed and edited the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-023-05042-3.

Correspondence and requests for materials should be addressed to Hugo J. Ladret.

Peer review information *Communications Biology* thanks Dario L. Ringach and the other anonymous reviewer(s) for their contribution to the peer review of this work. Primary handling editors: Enzo Tagliazucchi and Joao Valente. A peer review file is available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

4.4. Conclusion

In this article, we have studied how the (inverse) variance of oriented inputs affects V1. We have replicated some neural responses known in macaquesV1 [115], extending them to complex modulations of the dynamics of single neurons and populations. Using a prior-free machine-learning approach, we have probed for neural code in the population activity, uncovering two functional types of neurons that can encode either solely the sensory feature, or jointly the sensory feature and its variance. Based on cortical layer position and on a computational approach, we propose that the processing input variance in V1 is supported by recurrent connectivity between local cortical populations.

This directly ties the present article into the predictive processing framework introduced earlier in the manuscript (Equation 2.3), as the experiments here essentially mean consists in manipulating the variance Σ_u of a light pattern, in combination with its mean $g(v)$, respectively B_θ and θ in the article. While this means total access over the distribution of the input, two limitations on the predictive aspect of these recordings are missing. First, the distribution of internal priors were not controlled, for tractability's sake. This could have been done with an adaptation protocol, for example, by biasing the input over a given set of variables [69]. Second, we did not control whether the recorded neurons were encoding prediction errors or predictions. The former is rather simple to do, by inducing a mismatch between an unexpected sequence of patterns [162] (for example, an abrupt change of repeated orientations). The latter is more complex, as a major challenge in identifying them stems from their propensity to response like bottom-up activated neurons in various conditions [163].

Nonetheless, the present results provide an experimental validation of probabilistic processing in the brain, if not directly predictive (due to the anesthetized setup). This proves that there exists a mechanism by which the variance can weigh the activity of the V1, essentially linking back to Equation 2.26, in which the neurobiological implementation of Σ_p and Σ_u would be through recurrent connectivity in the brain, as predicted by the matrix form of predictive processing problems [31] (Figure 2.12). Further, the heterogeneity of recurrent connectivity, and the cross-orientation process between neurons as advanced in the final section of the article matches very well the notion that these specific interactions are competitive, inhibitory ones (Figure 2.13).

These discoveries also have very interesting ties to canonical models of orientation selectivity, namely to complex cells. As we have mentioned in the introduction of this chapter, these cells exhibit properties that stem from a pooling of simple cells [138], but that can also stem from recurrent heavy-computations [48]. Could it be that complex-cells are resilient cells ? A theoretical answer can be easily produced, as one could measure the phase invariance of the modelled resilient neurons in the article. An experimental answer would have required stimulating the neurons with a periodic pattern, and measuring their modulation rate [289, 257], which was alas not done here.

On the decoding side, our approach was based on a simplistic model of neural code, in the form of logistic regression. This was chosen for the biological plausibility
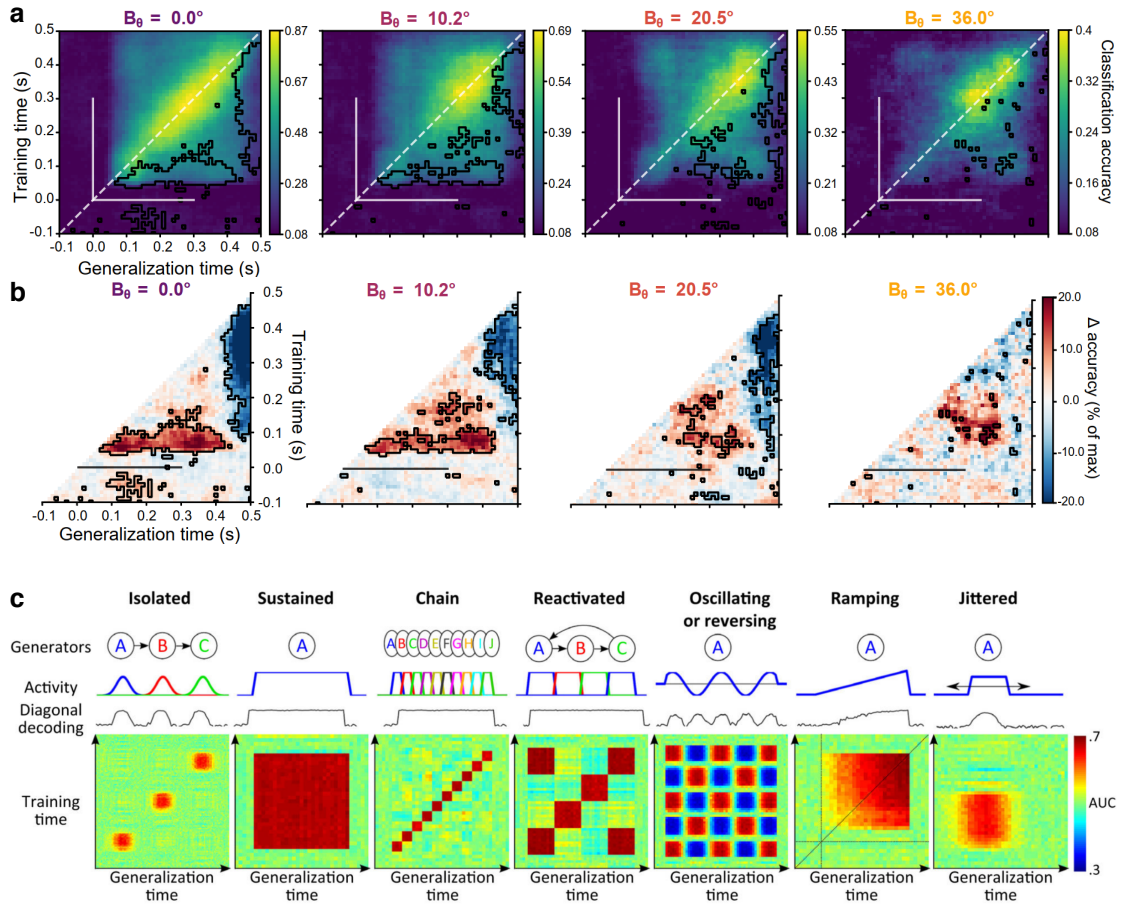
Figure 4.10. – Temporal generalization of the decoding. (a) Temporal generalization matrices for decoding θ at multiple B_θ. Filled white line represent onset of stimulation, and white dashed line elements where time of training and testing are identical. (b) Asymmetry of upper and lower matrices (around the diagonal). (c) Possible results from the temporal generalization process, adapted from [165].

of this read-out, as logistic regression essentially acting as a non-linear (sigmoidal) neuron, receiving information from the recorded spikes [27]. Although not shown here, we also probed for a number of additional algorithms, including support vector machines, deep feedforward networks, K-nearest neighbors and random decision trees, with no accuracy benefit for these increased interpretability and complexity cost [113]. Neural decoding strategies indeed span a wide spectrum of complexity, often incorporating highly intricate methodologies that leverage non-interpretable, multifaceted aspects of neural activity. A notable advanced example is the decoding of neural activity manifolds and their correlation to behavior [57], a subject garnering considerable attention in modern neuroscience research. A manifold, in the realm of neural decoding, refers to a high-dimensional space (but of lower dimension than raw data) constituted by the collective activity of a group of neurons. It represents a geometric framework where each point within this space corresponds to a unique state of neural activity. Decoding the manifold implies analyzing and interpreting the intricate structures and patterns within this high-dimensional space and correlating them to specific behaviors or cognitive states. This approach to deciphering neural activity seeks to uncover the underlying structures and relationships within the neural data, potentially revealing unprecedented insights into the dynamics of neural networks and their correlations to behavioral outcomes [272]. Manifold-based approach are enveloped in complexities and abstraction, but hold significant promise in advancing our understanding of the intricate interplay between neural activity and behavior.

In an earlier version of this manuscript, we placed more emphasis on the temporality of the neural code. For this, we turned to the method of temporal generalization, which consists in training the classifier on activity at a given time period, then testing it onto another one. For example, if a neural code is stable, one can train a classifier on early (post-stimulation) data, and observe good classification performance on late (post-stimulation) data. This allows to investigate whether the information carried by neuronal activity is consistent across different temporal phases following a stimulus. Further, specific shapes of the temporal generalization matrix can be (theoretically) tied to specific patterns of propagation of activity between neurons (Figure 4.10). In the present case, this was used to find a series of non-reversible neural codes, i.e. computations that do not generalize in both time directions which is a hallmark of a sequential, iterative computation [165]. More specifically, low variance stimulation contained maximally informative neural code early in the onset of the post-stimulation activity, which disappeared afterward. For high variance stimulation, the information was contained in later timestamps, which ultimately informs us on a time-dependent neural code (unsurprising, given single neurons dynamics shown in the article).

After having shown in this article a laminar (i.e., vertical) organization of variance computations, we turned to a new type of extracellular electrodes to extend those results. As interactions between neurons with different preferred orientation requires sampling of the orientation map, a horizontal probe is better suited to this task. As such, our results we replicated in awake macaque in Pieter Roelfsema's laboratory, using Utah arrays (Figure 4.11), which are matrices of electrodes, consisting of 128 recording sites in a 8x8 grid. Using the exact same type of stimulations (MotionClouds,
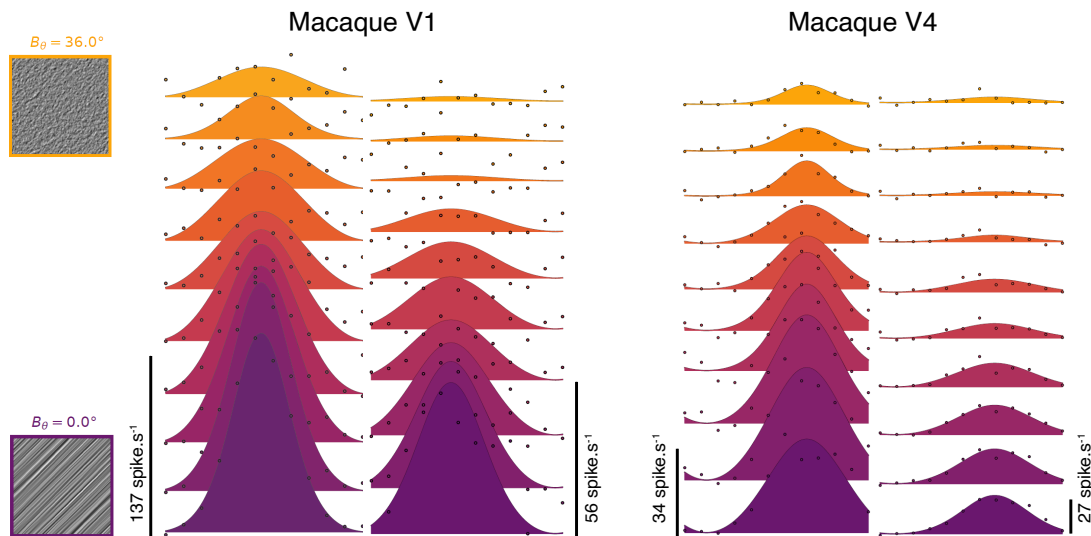
Figure 4.11. – Orientation variance modulations in macaque V1 and V4. Stimulation protocol is similar to the one presented in this chapter's article. These results will be further elaborated upon in chapter 5.

rigit drift orthogonal to median orientation), we replicated similar neuronal behavior from anesthetized cats to awake primates, but also to extend our results into extrastriate areas, namely in V4. Interestingly, this not only holds for the single isolated neurons as shown in Figure 4.11, but also for the response of neurons that are anatomically grouped. The specific advantages of this method of sampling are showcased in chapter 5, in which we infer functional connectivity directly into this data.

5. Mapping Neural Interactions in V1: A Graph-Based Perspective

"I won't be round this old town,
anymore for a long, long time,
gonna hit the road and start looking for the end of that long white line"
Sturgill Simpson, Long White Line, 2014

Summary

5.1. Introduction: Towards Mesoscale Recordings in V1

Unlike the previous chapters in this manuscript, chapter 5 does not feature an already published or submitted article. This difference is due to the chapter focusing on a new set of experimental data, collected in Dr. Pieter Roelfsema's lab by Dr. Paolo Papale. These unique results stem from recordings in a single awake vigil macaque, which was initially implanted with Utah arrays for a different study. We were fortunate to have Drs. Roelfsema and Papale's support, allowing us to perform recordings linked to Motion Clouds in this primate model, building upon the discoveries highlighted in chapter 4. However, it's important to note that standard practices for publishing require data from at least two primates. The decision to further collaborate and include a second primate in this study was contingent on the preliminary results obtained from the first primate. While this process is underway, it is expected that the resulting data will be available after the completion of this manuscript.

Therefore, chapter 5 will be structured similarly to an article, but will not include one. It will also be (very) short, compared to other chapters, as it mainly aims to reinforce significant concepts introduced earlier, although the current stage of the data is still too preliminary for a more extensive discussion.

Indeed, we will here build on an implicit concept of the manuscript's introduction: the notion that processing variance in low-level sensory cortices (i.e., V1) relates to learning the structure of the variance in the environment. This proves advantageous for adaptive coding properties, and for redundancy elimination [18]. On a methodological point of view, extension from anesthetized cats to awake primates is not trivial (see also the conclusion of this manuscript). It eliminates the possible confounding factor of anesthesia, and enables comprehensive neurobiological sampling of recurrent activity with Utah arrays (Figure 4.8).

This is a significant step forward from the approach in chapter 4, where we relied on interpolating through data from different sources and computational approximations to understand neural dynamics. In that sense, Utah arrays are advantageous by allowing for the sampling of multiple, distinct sites across the orientation map. However, Utah arrays present another challenge due to their higher impedance (inverse of the resistance), meaning they effectively record from a larger area around their implantation site. This larger recording area makes it impractical to perform the precise spike sorting and manual curation, as done in chapter 4.

Instead, our approach will focus on Multi-Unit Activity (MUA). Recording MUA is akin to recording the aggregated activity of nearby neurons, and has been shown to be a good proxy for single unit recordings in structured topological maps [32, 226, 181] like V1. Certain limitations apply: the temporal nature of the signal is less accurate, and if a given recording site happens to fall in the "pinwheel" of an orientation map (Figure 2.5), it is likely that competing interactions are going to be part of the signal recorded. Despite the different methodological constraints imposed by the use of Utah arrays, this chapter will extend on the notion that that computations of variance in neural signals are influenced by interactions among neurons, which are distributed, rather than concentrated, when variance increases.

5.2. Methods: Graphs of Neural Activity

As described above, recordings were carried in a single awake, vigil macaque, with Utah arrays implanted in V1 and V4. Extrastriate data will not be part of the present manuscript, but do show similar modulation of orientation selectivity by orientation variance as V1 (as shown in Figure 4.5). This consistency across areas does suggest potential avenues for future research, discussed in the conclusion of the manuscript.

To briefly detail the experimental setup, the framework was similar to that of chapter 4, modified with adaptation for primate recordings. Here, the macaque was engaged in a passive fixation task, which involved the presentation of Motion Clouds, shown for 300 ms and interleaved with mean luminance screen for 150ms. Median orientation θ was varied between $0; \pi$ in 12 even steps, and orientation variance B_θ in 10 steps between $\pi/30; \pi/3$. Other parameters were adapted from chapter 4 to match macaque V1 preference, namely a spatial frequency of 1.2 cycles per degree, and a drifting speed of 3 cycles per second [251]. Stimuli were drifting in either direction, orthogonally to median orientation, and were averaged across direction. The macaque was headfixed,

and observed a 1024 by 768 pixels screen from a distance of 47.5cm. All stimuli were shown at least 40 times.
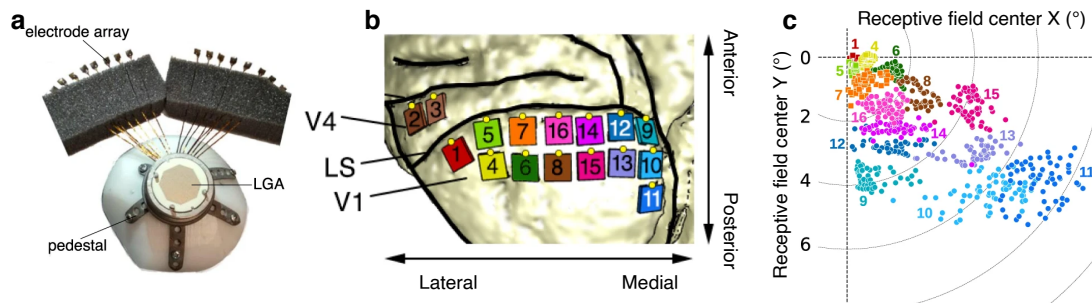
Figure 5.1. – Typical setup of Utah array recording in macaque visual cortex, similar but not identical to the one used in the present chapter. (a) Example of a macaque implant, consisting of 1024 channels encased in a titanium cranial pedestal. (b) Typical layout of the arrays. Here, "only" 8 arrays, i.e. 512 channels, are implanted in V1. (c) Typical eccentricity of the receptive field, as recorded here. Reproduced from [55].

MUA was recorded in 8 Utah arrays, consisting of 8x8 electrodes, hence totaling a number of 512 electrode sites (Figure 5.1). Given that the implantation of the chamber was done for another set of experiments, it was necessary to remove certain recording sites that were no longer yielding good signal [306]. A signal-to-noise ratio was measured, based on evoked activity over resting (300 ms before evoked) baseline. This ratio was measured as the peak activity subtracted by mean resting activity, and normalized by the standard deviation of that resting activity [55]. Channels with a signal-to-noise ratio below 1 were removed from further analysis, leaving 384 electrodes. As in chapter 4, the circular variance served as a criterion for exclusion of untuned neurons. A threshold was set at 0.85, removing untuned electrodes and leaving a final 188 electrodes for analysis.

We thus aimed to analyze how information transfer between each of the remaining 188 electrodes is affected by input variance. There are numerous ways (237, to be exact [58]) to measure pairwise interactions between sets of time-based data, and each method comes with its unique benefits and limitations. Here, we decided to measure covariance, which is one of the simplest and most intuitive methods available, and represents how much two electrodes evolve together. This covariance matrix is similar to the matrix form of variance, but extended to multiple variables, as used in the matrix formulation of predictive processing (Equation 2.28). Given a pair of electrodes X and Y, the covariance at a single timepoint is:

$$\text{Cov}(X, Y) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1} \tag{5.1}$$

where X_i and Y_i are the individual values of the signal in electrodes X and Y for a given trial i, $\bar{X}$ and $\bar{Y}$ are the mean values accross trials, and n is the number of trials

(here, 40).

Not only is covariance straightforward to compute and interpret, but its matrix inverse is also useful, as it is the inverse variance matrix, which is the central focus of this thesis. When dealing with large datasets, like ours with 188 variables (electrodes) but only 40 observations (trials), directly inverting the covariance matrix can lead to instability due to the disproportionate ratio of variables to observations. To address this, we employ a technique known as a "shrinkage estimator". This method provides a more reliable estimation of the inverse covariance matrix, making it feasible for inversion. Here, we used the common Ledoit-Wolf estimator, which given a matrix M:

$$\hat{\Sigma}_{LW} = \alpha \cdot F + (1 - \alpha) \cdot M \tag{5.2}$$

where $\hat{\Sigma}_{LW}$ is the Ledoit-Wolf shrinked matrix, α controls a shrinkage intensity, F is the target matrix (the identity matrix scaled by the average variance) and M is our sample covariance matrix. Here, α is automatically chosen to minimize the mean squared error of the reconstruction [188] [1]. The inverse variance matrix (or precision matrix), is thus $\hat{\Sigma}_{LW}^{-1}$.

A property of covariance matrices, and thus of inverse variance matrices, is their symmetry: the interaction between the electrode X and Y is the same as the interaction between electrode Y and X. As a result, the interactions we observe between electrodes in this context are not directional; they only have varying degrees of intensity, or weight. While directional metrics might be useful for uncovering how information is transferred between neurons, potentially exposing interactions that are unique to specific orientations, this is being the scope of the current research.

We visualized these matrices as graphs, which are composed of nodes (representing electrodes) and edges (indicating covariance or inverse variance). This graphical representation makes it easier to comprehend the data, as the matrices in their raw form are not readily interpretable. Additionally, this approach offers a range of mathematical tools to analyze network variations. Here, the graphs are non-directional (because of the covariance or precision symmetry), cyclic (because they form closed interactions loop) and weighted (by the co-variance/inverse variance). To manage complexity and minimize noise, we binarized these graphs. This means we disregarded edges below a certain threshold, as a $188^2 = 35344$ edges-wide graph would be too unstable for analysis. We retained only those edges that ranked above the 75th percentile in weight distribution, a common practice in graph theory analyses [82], This narrowed the focus on the most significant interactions within the network, at the cost of full network representation.

The specific nature of the graphs we've created here limits the range of questions we can explore regarding their topology. To navigate this, we've selected six key

---

1. A year before the publication of this article, these authors released a related article entitled "Honey, I Shrunk the Sample Covariance Matrix" [187]. For some reason, the author of the present manuscript found this amusing enough to be note-worthy, and even considered at one point that it would make an excellent title for this chapter. This was eventually decided against, and regarded as a sign that it might be time to stop working on inverse variance-weighting and graduate.

metrics to analyze: clustering, centrality, assortativity, neighboring connectivity, small-worldness, and the Wiener Index [206, 42]. Mathematical details are provided hereafter, but are not necessary for the comprehension of the results. Rather, an intuitive description is provided with each equation, and a graphical representation is shown in Figure 5.2.
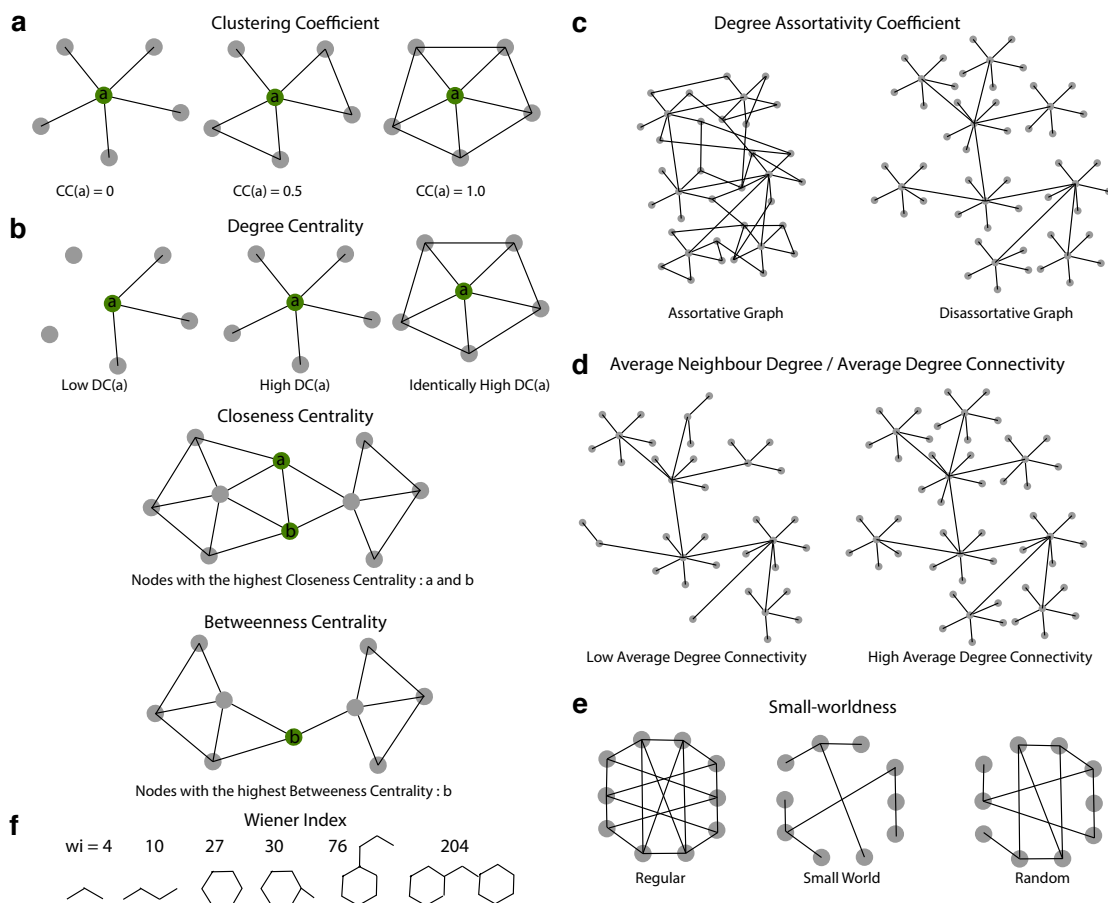
Figure 5.2. – Graph metrics, presented in identical order as in the text below. (a) Clustering Coefficient, (b) Centrality metrics, (c) Degree Assortativity Coefficient, (d) Average Connectivity Metrics, (e) Small-worldness index, (f) wiener Index, represented onto common chemical compounds (delocalized electron cycles not shown).

The first metric used is the clustering coefficient, a measure of the degree to which nodes in a graph tend to cluster together. For example, in the context of a social graph, it helps understand the extent to which friends of a given person are also friends between each other. Here, it can be seen as a direct measure of the global interconnectedness of the neural network being recorded. The equation for the clustering coefficient C_i of a node (electrode) i in a graph is:

$$C_i = \frac{2T(i)}{k_i(k_i - 1)} \tag{5.3}$$

where $T(i)$ is the number of triangles through the node i (i.e., the number of connec-
tions that exist among the immediate neighbors of the node i) and k_i is the degree
of node i (i.e., the number of edges connected to node i). C_i ranges from 0 to 1, with
higher values meaning a greater degree of clustering. Here, all node-wise metrics were
reported as a distribution of values across multiple input θ, for each B_θ.

   A second metric, or rather, class of metric, is the measure of centrality. Multiple
sub-metrics are available here, each with their own importance. We measured three,
which are:
   — Degree Centrality: The simplest form of centrality. For a node in a graph, degree
     centrality is simply the count of how many connections (edges) it has. Nodes
     with higher degree of centrality are typically more influential or important in
     a network because they have more connections. Thus, the higher the overall
     Degree Centrality of a graph is, the more connected the graph is. Following the
     equation above, it is thus simply measured as k_i.
   — Closeness Centrality: This centrality metric focuses on how close a node is to
     all other nodes in the network. It is defined as the reciprocal of the sum of the
     shortest path distances from a node to all other nodes in the network. A higher
     closeness centrality indicates that a node can spread information to all other
     nodes in the network through less synapses. It is often correlated with the metric
     above, but not necessarily (see Figure 5.2 for example). For a given node v, the
     closeness centrality $C_C(v)$ is:

$$C_C(v) = \frac{1}{\sum_{u \neq v} d(v, u)} \tag{5.4}$$

     where $d(v, u)$ is the shortest-path distance between nodes v and u, and the sum
     is taken over all nodes u in the graph except v itself.
   — Betweenness Centrality: This centrality metric quantifies the number of times
     a node acts as a bridge along the shortest path between two other nodes. It
     captures the degree to which a node lies on paths between others, indicating its
     potential for control over information flow in the network. Nodes with higher
     betweenness centrality can have significant influence within a network, by virtue
     of their ability to gate message passing between other nodes. For a given node v,
     the betweenness centrality $C_B(v)$ is calculated as:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{5.5}$$

     where σ_{st} is the total number of shortest paths from node s to node t, $\sigma_{st}(v)$ is
     the number of those paths that pass through v. The sum is computed over all
     pairs of nodes s and t in the graph, where $s \neq t \neq v$.

   A third metric is the Degree Assortativity Coefficient. It measures the tendency of
nodes to connect to other nodes which have similar degrees. In other terms, it reflects
whether high-degree nodes (nodes with many connections) tend to be connected to
other high-degree nodes, and similarly for low-degree nodes. The Degree Assortativity

Coefficient r can be calculated using the Pearson correlation coefficient for degree-degree pairs across all edges in the network. The formula is as follows:

$$r = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sigma_q^2} \tag{5.6}$$

where e_{jk} is the proportion of edges in the network that connect a node of degree j to a node of degree k, q_j is the proportion of ends of edges that are attached to nodes of degree j (also known as the "normalized degree distribution") and σ_q^2 is the variance of the distribution q. The assortativity coefficient r ranges from -1 to 1. A value of 1 indicates perfect assortative mixing patterns, 0 indicates non-assortative mixing, and -1 indicates perfect disassortative mixing. Understanding the degree assortativity of a network is essential in analyzing the robustness and the dynamics of information or disease spread within the network. Thus, if high-degree nodes tend to connect with low-degree nodes, the network exhibits negative degree assortativity, as shown here.

Fourth are degree metrics, namely, the Average Neighbor Degree and the Average Degree Connectivity. The Average Neighbor Degree is a measure for each node in a network, that is the average degree of its neighboring nodes. This measure is useful for understanding the tendency of nodes to connect to others that are similarly well-connected or not; much like the Degree Assortativity Coefficient. For instance, in a network with high degree correlation (assortativity), high-degree nodes tend to be connected to other high-degree nodes. For a node v with degree k, the Average Neighbor Degree AND(v) is given by:

$$\text{AND}(v) = \frac{1}{k} \sum_{u \in N(v)} \deg(u) \tag{5.7}$$

where $N(v)$ denotes the set of neighbors of v, and $\deg(u)$ is the degree of a neighbor u. The Average Degree Connectivity, on the other hand, is a network wide metric that measures the average degree of the neighbors of nodes with a given degree. As such, it is often correlated with the AND. The Average Degree Connectivity for nodes of degree k in a network, denoted as ADC(k), is defined as:

$$\text{ADC}(k) = \frac{1}{N_k} \sum_{v:\deg(v)=k} \text{AND}(v) \tag{5.8}$$

where N_k is the number of nodes with degree k in the network. Their interpretation, similar to the Assortativity Coefficient, gives insight into whether the network is assortative or disassortative (Figure 5.2).

A fifth metric is the degree of small-worldness, which quantifies a network in which clustering coefficient is high, while maintaining short average path length. In neural networks, this results from the balance between minimizing the resource cost and maximizing the flow of information among the network components. In brain-wide networks, the metabolic cost between neighboring neurons is much lower than that of distant neurons, and thus the brain behaves a small-world network [191] to increase

efficiency, as described in the introduction of this manuscript [18]. A direct measure of small-worldness compares the current graph with a random network [336]. This measure of small-worldness quantifies the balance between local clustering and global reach in a network. Specifically, small-world networks have significantly higher clustering than random graphs but similar average path lengths. Thus, the small-worldness S of a network can be quantified as follows:

$$S = \frac{C/C_{random}}{L/L_{random}} \tag{5.9}$$

where C is the average clustering coefficient of the network, L is the average shortest path length of the network, C_{random} and L_{random} are the average clustering coefficient and average shortest path length, respectively, of an equivalent random graph. Typically, a network is considered to exhibit small-world properties if $S > 1$, indicating that it has higher clustering than a random graph while maintaining a comparable average shortest path length.

Finally, a sixth metric is the Wiener Index, one of the oldest topological indices [70] that is used primarily in chemical graph theory (see Figure 5.2). It is a measure of the compactness of a network and is closely related to the small-worldness and efficiency of the network. The Wiener Index W of a network is defined as:

$$W = \sum_{\{i,j\} \subseteq G} d(i, j) \tag{5.10}$$

which is the sum of the shortest path lengths between all pairs of nodes in the graph. In this context, G is the set of nodes in the network, and $d(i, j)$ represents the shortest path between nodes i and j.

5.3. Results: Modulations of Connectivity Patterns by Orientation Variance

Before diving into graph metrics, let us first assess direct V1-based metrics. Namely, we will first observe whether the variance-tuning curves shown in chapter 4 are also present in primate V1. In this part of our study, the reliance on MUA, rather than single-neuron activity, means we lose the granularity of precise spike timing. There are however notable parallels in how both MUA and single neuron activities in terms of orientation tuning [181].

This holds true here, as the variance-tuning functions reveal similar patterns of activity modulation across neurons as shown in chapter 4. Despite the shift in scale from single neurons to V1, we observe a comparable distribution of heterogeneous modulations in neuronal activity, as shown in Figure 5.3. This validates the consistency of our findings across different levels of neural activity and species. This, in turns, allows us to extrapolate (with caution) some similarities between the precise single-neuron mechanism of chapter 4 and the graph metrics of this chapter.

Figure 5.3. – Variance tuning functions in primate V1. (a) Modulation of tuning curve for two example MUA, reminiscent of resilient/vulnerable neurons described in chapter 4. (b) Variance-tuning functions of these two examples MUA, fitted with a Naka-Rushton function (black line). (c) Population map of variance-tuning functions, with dots representing $B_{\theta 50}$, the changepoint of the function (see chapter 4). (d) Population map of variance-tuning functions, as with (c).

While timing modulations are present (data not shown), in practice, these were on the order of tens of milliseconds in single neurons (i.e., about one τ), and thus, it would be hard to extrapolate these in MUA due to smoothing across multiple neurons. Given these constraints, our analysis at the single-electrode level is largely limited to tuning metrics. Hence, we turn our focus to mesoscale measurements, which offer a broader view of neural activity while still providing valuable insights. An example of such a mesoscale measurement is neural decoding, as we explored in chapter 4. Decoding would have been a particularly interesting approach in this context, especially since the activity is recorded in physical simultaneity, as opposed to virtually concatenated in chapter 4. Given similiarities in tuning, we might hypothesize that the decoding results in primate V1 would mirror those found in cat. This assumption is based on the consistent tuning characteristics observed across different animal models, suggesting a fundamental similarity in how V1 processes visual information.

However, here, we tried to branch away from simply reproducing chapter 4 results, and opted to investigate the interactions between neurons, as opposed to decoding an emergent neural code. As detailed in the methods section, to explore this, we utilized covariance matrices and their inverses, the inverse variance matrices, as our primary analytical tools. For the 188 electrodes, such matrices are shown in Figure 5.4. These represent the interactions between neurons, and characterize the variation of recurrent message passing between neurons. Here, there are measured for a single input θ, which was chosen as the one with most MUA tuned onto. We reasoned that this would be the orientation for which the representation would be the most accurate, but we will then average metrics across orientations in later sections.

Interpreting connectivity differences solely from the matrices can be challenging, and one often sees what they wish to be true. Even employing ratios or advanced metrics like the Laplacian or matrix-norm doesn't significantly enhance our understanding of these interactions. To mitigate this, we later broaden our approach to average across different orientations, aiming to capture a more comprehensive view of neuronal interactions across the visual spectrum. This expanded analysis is part of our ongoing effort to better understand the intricate web of connections and influences within neural networks.

Thus, it is better to embed these matrices into graphs, which offers better representations and better metrics. These graphs can be plotted onto a tuning ring, where each node's preferred orientation corresponds to a specific position on the circle. This is a neurobiological realization of the computational "ring" model we presented in chapter 4. By positioning neurons according to their orientation preferences, we create a visual map that reflects the inherent structure of neural preferences and interactions in the visual cortex. This is shown in Figure 5.5, which shows a clear variation on the condition of interactions.

This translates into an immediate, non-scientific conclusion: increasing the variability of input tends to complexify the graph's appearance. There is little to interpret with these representations, as the graph used here are fully connected (due to being based on covariance matrices). How does this translate into pratical interactions? A more effective method for visualizing these graphs is through a force-directed algorithm,
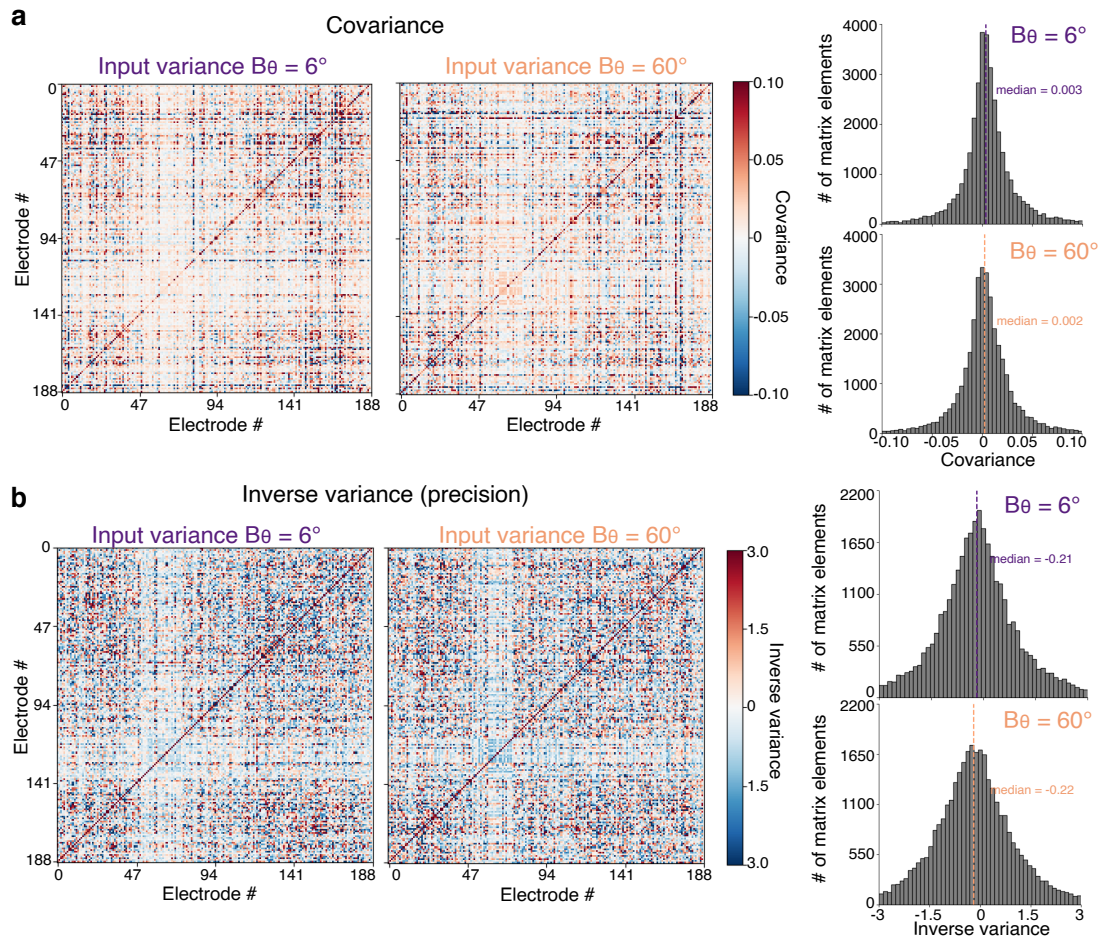
Figure 5.4. – Inverse variance-weighting in primate V1. (a) Covariance matrices for lowest and highest variances, for a single orientation. The order of the elements, from 0 to 188, follows the preferred orientation at each electrode. Distribution of the covariances values with median value is shown in the right. (b) Same as (a), but for inverse variance (precision) matrices.
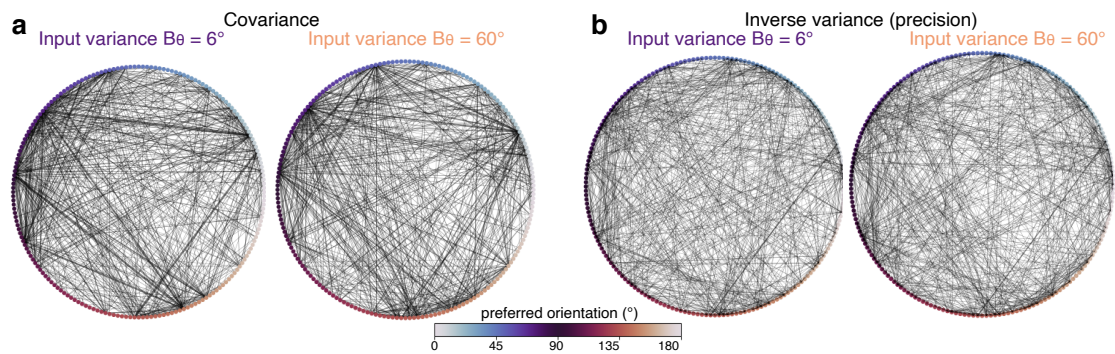
Figure 5.5. – Circular graphs of inverse variance-weighting in primate V1. (a) Graphs of covariance, with units sorted by their preferred orientations, for lowest and highest variances. (b) Same as (a), but for inverse variance.

specifically the Fruchterman-Reingold force-directed algorithm, often referred to as a spring layout [101]. This attempts to maintain a certain distance between each node, essentially pushing them apart, but holding them together through their interactions. For instance, positive inverse variance-weighting draws the nodes closer, whereas negative weighting drives them even further apart. This represents the graph in a manner where all the connections (edges) are relatively equal in length and overlap as little as possible, creating a graph that is more straightforward to comprehend. This is shown in Figure 5.6.
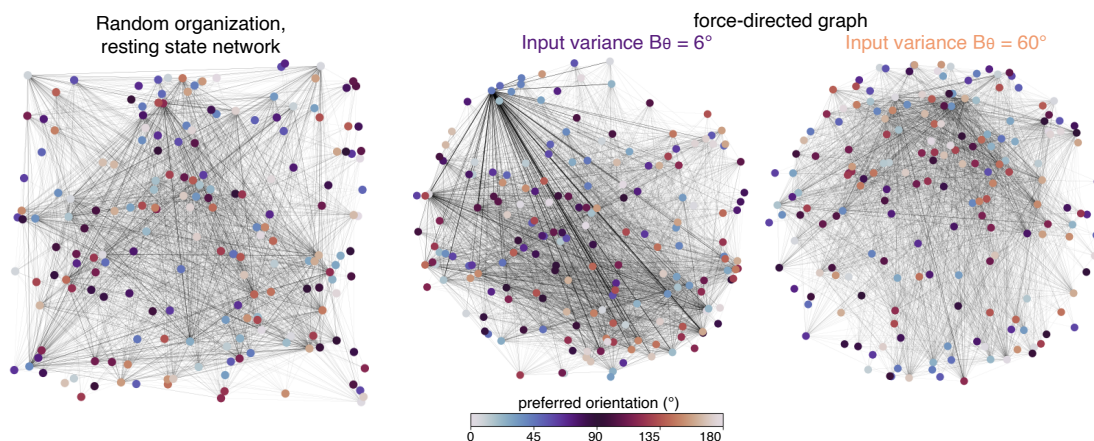
Figure 5.6. – Force-directed graphs of inverse variance-weighting in primate V1, initialized from random positions (left), for lowest (middle) and highest variance (right).

This is now steering us towards a clearer interpretation of the graph's dynamics. Through these representations, it becomes evident that with maximum input variance, the network's structure transitions from a dense, small-world configuration (concept introduced in Figure 5.2) to a more dispersed, globally distributed network. This observation aligns with the concepts introduced in chapter 4, namely the idea that computations within the network become more iterative and distributed when dealing with high variance, in contrast to being more concentrated and feedforward-based for inputs with low variance.

To properly quantify this phenomenon, we will now turn to the metrics that have been described in the methods section above (Figure 5.7). In the order in which they have shown in Figure 5.2, these metrics paint a converging picture:

— Clustering Coefficient increases as a function of input variance (Mann-Whitney U test between $B_\theta = 0°$ and $60°$: $U = 0.0, p < 0.0001$. Linear regression : $y = 0.004x + 0.68$, $p = 0.01$). Intuitively, this means that the neural activity in V1 is becoming more interconnected, as opposed to separated into uncorrelated sub-networks. This implies a shift from isolated processing to more integrated, recurrent neural activity, where neurons are not just operating independently, but are increasingly interlinked.

— Degree Centrality ($U = 1968.5, p < 0.0001$. lin. reg. : $y = 0.006x + 0.65$, $p = 0.005$)

and Closeness Centrality ($U = 1994.0, p < 0.0001$. lin. reg. : $y = 0.003x + 0.74$, $p = 0.006$) increase as a function of input variance. This gives a similar interpretation as with the Clustering Coefficient, that of a network whose units are processing information in a more distributed fashion. Betweenness Centrality, however, is decreasing with input variance, ($U = 19555.0, p < 0.0001$. lin. reg. : $y = -0.00003x + 0.001$, $p = 0.005$), suggesting that fewer nodes are acting as critical bridges within the network. This could mean that the network is becoming less reliant on given neural pathways, and distributes its processing load more evenly across the network.

— Degree Assortativity Coefficient is negative for all input variance, but still increases with input variance (lin. reg. : $y = -0.005x - 0.05$, $p = 0.0002$). This means that a dissasortative neural network is progressively reorganizing into a less-tightly organized assortative neural network. However, as this coefficient absolute value's does increase with variance, it indicates a shift towards a network where nodes are more likely to connect with similar nodes. This transition suggests a move from a more hierarchical or specialized network to a more homogenized one.

— Average Neighbour Degree is increasing with input variance ($U = 0.0, p < 0.0001$. lin. reg. : $y = 1.05x + 124$, $p = 0.006$). This implies that that neurons are not only increasing their connections but are also tending to connect with other neurons that are similarly highly connected, hinting at a more uniformly interconnected network, where all neurons are part of equally significant ensembles of interactions.

— Small-worldness is decreasing with variance (lin. reg. : $y = -0.004x + 1.06$, $p = 0.001$). This marks a departure from the dominant default-mode of connectivity in neural networks [206], and as with Average Neighbour and Average Degree Connectivity, V1 is becoming less tightly organized and distributed.

— Wiener Index is decreasing with variance (lin. reg. : $y = -119x + 23820.5$, $p = 0.004$). A decrease in this index is intriguing, as it suggests that the overall path lengths within the network are reducing. This could mean that despite the increase in distributed processing, the network is somehow becoming more efficient in terms of signal transmission distances, which could be due to the reorganization of connections in the network. One implication of this is that the seeming disorganization of the network is not so much a disorganization as it is a planned reorganization, into an equally efficient but topologically different neural network.

Overall, these findings are very exciting, as they all converge towards a similar observation: with increased input variance, V1 is actively being reorganized from a tightly segregated topology into a delocalized, recurrent processing neural network. How these metrics evolve through time could also align with the idea of dynamical processing of orientation variance, as exposed in chapter 4, and form a promising avenue of research.
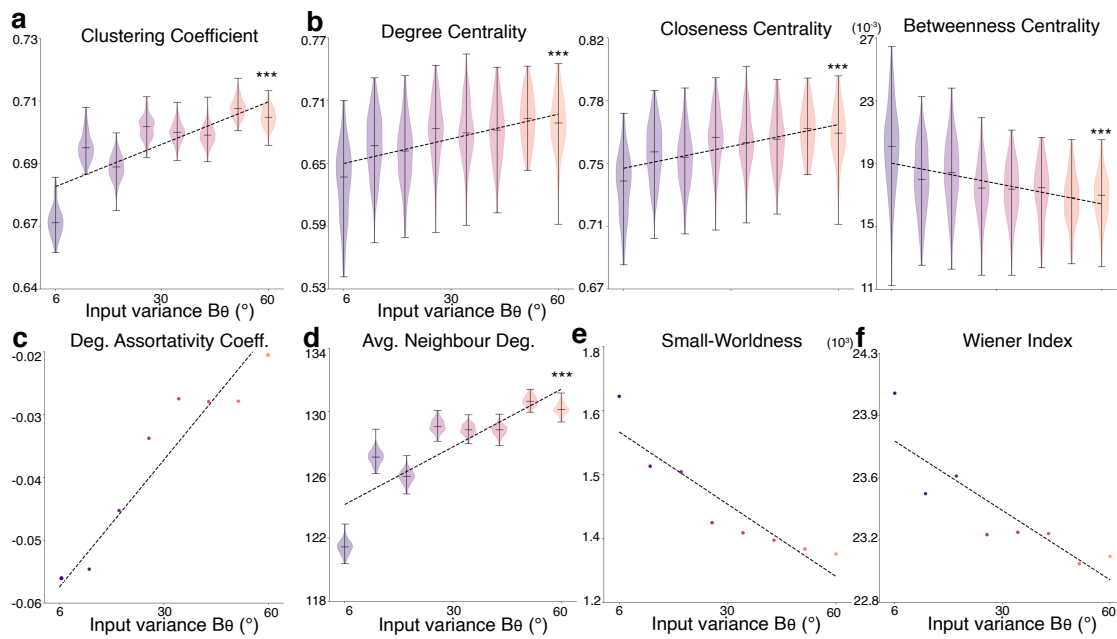
Figure 5.7. – Quantification of graphs. (a) Clustering Coefficient, showed as a vio-
linplot with 5th, 95th quartiles and median values. Linear regression is
shown as a dashed black line. Significance of the difference between
$B_\theta = 0°$ and 60° is indicated above the last boxplot (Mann-Whitney U test,
$* * *; p < 0.001, * *; p < 0.01$). (b) Centrality metrics. (c) Degree Assorta-
tivity Coefficent. (d) Average Neighbour Degree. (e) Small-Worldness. (f)
Wiener Index.

5.4. Conclusion: A Predictive Coding Perspective

Graph-based approaches to neuroscience often describes the brain as a small-world network [191], which is a general principle of any system that balances efficiency and minimization of operating cost [298]. In the context of our manuscript, this means a series of networks that function in tandem, and are coordinated by broader neural interactions. Although the leap is theoretical, these can also be likened to a 'canonical microcircuit' model, where the brain processes sensory information (like orientation) and puts collaboration between adjacent circuits into play to form a comprehensive and integrated model of the environment.

Our research, particularly in Chapter 4, delved into the functional organization within a single cortical area, highlighting how single neurons and populations adapts to different variance in sensory input. We theorized an increase in recurrent neural activity in response to higher input variances. Here, in awake vigil macaques, we show that is experimentally the case, using a graph theory approach that shows decentralization of the neural network as a function of the sensory input's variance. This aligns with our initial theory that the brain's segregated and linear processes evolve into more interconnected, decentralized operations involving multiple neurons, when faced with complex stimulus. This also implies that experiments based on stimulating V1 with drifting gratings are missing a crucial key component of the brain, that is, its network capacity to re-organize in face of sensory variance to adjust to an uncertain environment.

The implications of these findings for predictive coding are significant. Let us now focus on the notion of variance, rather than inverse variance, as this will make more intuitive sense based on a Motion Clouds definition. If input variance increases in the visual input, patterns of activity across multiple orientation-tuned units in V1 also increase in variance (as also shown in chapter 4). This decentralization of the message passing between neuron implies that the distribution of neural activity, in orientation space, is a function of the distribution of input orientation. As posited by predictive coding, lower-level (inverse) variance distributions in sensory areas are essential for learning the statistical variance in the environment. Essentially, the brain uses these lower-level distributions to fine-tune its predictions about the inherent variance in both the environment, but also in the sensors used to sample said environment.

To validate this notion, one could imagine showing repeated high variance input to a macaque, and see how these response changes from a V1 that is used to process heterogeneous variance. One could then expect systematically broader patterns of activity in V1. Another way to explore these limits is to implement them into a predictive coding model, and see how this model behaves in the face of changing input variance. With the theoretical and mathematical framework already in place in the thesis' introduction, it would be rather straightforward to develop a neural network that implements predictive processing. This network could then be tasked with a simple machine learning classification task, for instance, solving the Modified National Institute of Standards and Technology database, a list of handwritten digits between 0 and 9. This task is trivial, and often the first one to be done on a newly

developed neural network. Following this tradition, one could embed these digits into Motion Clouds, thereby manipulating the orientation distributions of these inputs and see whether similarities in the learned inverse variance matrix follow those observed in macaque V1.

6. Beyond V1: Variance and Thalamo-Cortical Loops

"Welcome back my friends
to the show that never ends.
We're so glad you could attend,
Come inside! Come inside!"
Emerson, Lake and Palmer, Karn Evil 9, 1973

Summary

6.1. Introduction: The Functional Anatomy of the Pulvinar

After having had the pleasure of immersing ourselves in more than a hundred of pages of V1-centric research, it now seems only natural to take a break away from the striate cortex. As alluded to in the thesis's introduction, V1 takes part in parallel computations distributed both in extrastriate regions and in subcortical nuclei. These interactions are crucial for inverse variance processing, and now call for our attention for a proper description of a "multiscale" computational system. One particularly intriguing facet of these interactions is the role that the pulvinar nucleus plays. This thalamic nucleus operates as a sophisticated "routing" system, modulating the information flow between V1 and extrastriate regions [46]. Given the interest we took in the modulation of integration between multiple cortical regions by the inverse variance, the presence of a thalamic nucleus interconnected with this entire network is undeniably a brain region worthy of deeper exploration.

Before detailing more the involvement of pulvinar in predictive coding, we shall briefly introduce its anatomy and functions. The first article of this chapter consists of

a review of the role of pulvinar in this domain, aiming to introduce similar notions, and hence it would be redundant to go into too many details here. It should be noted that, in the context of this thesis, some theories suggest that the pulvinar might be the primary locus of inverse variance weighting in the brain [153], while others suggest it only plays a part in this role [211].

Anatomically, the pulvinar lies over the dorsolateral posterior thalamus, running alongside the edge of the LGN. In primates and in cats, the pulvinar is the largest thalamic nucleus, having increased in size throughout mammalian evolution with the visual cortex [139]. Given its size, the pulvinar can be further subdivided in six sub-regions in primates (human included) or three main regions in the cat [13], which, out of scope, will be here simplified as a single conceptual entity in this introduction (but see the first article for further details). The pulvinar establishes extensive reciprocal connections with all visual cortical areas [148], often being conceptualized as the "seventh cortical layer" [282]. In a striking difference with the LGN, which receives its main visual inputs from the retina, the pulvinar receives its visual inputs principally from the layers 5 and 6 of the visual cortex. This is reflected in the response characteristics of its neurons, which remarkably resemble those found in the visual cortex in terms of feature selectivity [205], but also in their implementation of selective attention [338, 332]. This allows pulvinar neurons to effectively suppress responses to visual distractors [258] by integrating the visual signals from multiple cortical areas and outputting modulatory signals to layers IV and I of the visual cortex. Overall, the pulvinar's connectivity gives it an integrate-and-modulate role within the visual hierarchy. This description is especially relevant when considering the visual system as a hierarchical generative Bayesian model within the predictive coding framework (Equation 2.28).

As put forward by Kanai et al. [153], the pulvinar's engagement in attentional regulation and its output connectivity to superficial cortical layers is an ideal match for a role in inverse variance modulation of the visual hierarchy, which also fits the long timescales of multi-areas integrations. Experimental results have shown that inactivation of the pulvinar drastically suppresses responses of superficial V1 neurons to sensory inputs, while excitation increases the responsiveness of neighbouring superficial units with overlapping receptive fields [252], which is similar to a gain-control mechanism that implements inverse variance-weighted modulation of the cortex by the pulvinar. Whilst it is the biggest thalamic nucleus in primates, the pulvinar is nonetheless an order of magnitude smaller than the entire visual cortex. As estimating inverse variance requires computing the co-variance between the all N prediction or prediction errors units, this is an N^2 computation [211]. Given the cost of such computation, it seems more likely that the pulvinar computes an approximation of the inverse variance matrix, and then modulates parts of the superficial layers independently. In parallel, these units will then use their local horizontal connectivity to further integrate and refine the computation of the local variance [153, 20], as proposed in chapter 4 [178]. Evolutionarily, this hypothesis is very interesting, as some species, such as birds, have much lower homologous "cortical" (i.e. pallium) neuron count, and seem to lack a functional equivalent of the pulvinar. In a thesis committee

discussion with Dr. Paul Cisek, we hypothesized that as the number of neurons grow exponentially throughout evolution, computing variance through recurrence alone (as proposed in chapter 4), becomes either prohibitively slow or computationally expensive. This thus requires a dedicated central part of the neural network to perform a dynamical approximation of the inverse variance matrix, which would hence be the role assigned to the pulvinar.
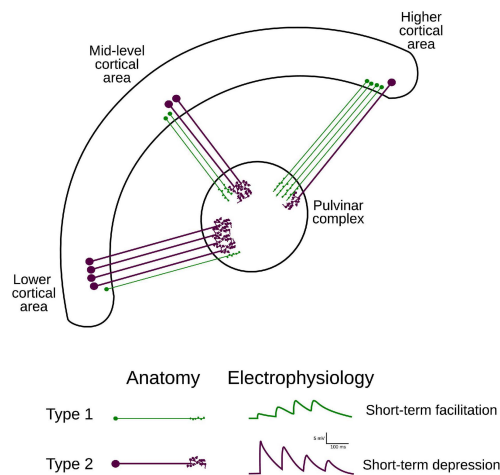
Figure 6.1. – Anatomical substrates of pulvinar gain control, from [61]. Note that the ratio of Type 1 over Type 2 synapses changes as a function of the visual hierarchy.

Overall, this depicts a dynamical gain control role for the pulvinar, much like the gain control role we proposed in chapter 4. How is such dynamical gain control in the visual cortical areas implemented ? Guillery and Sherman [121] characterized synapses associated to thalamic first-order relays, which transfer information about the world to primary cortical areas (in vision, the LGN). Two types of inputs, modulators and drivers (type 1 and type 2 projections, respectively), have been identified on the basis of their axon terminals' morphology and their electrophysiological characteristics [280]. Drivers and modulators can also be defined by other attributes, such as the input-output transfer of the neuronal response profile. Modulators effects are distinguished as either multiplicative or divisive effects (nonlinear gain control), while drivers act by either additive or subtractive changes (linear gain control) [2, 285]. In V1, it is generally accepted that pulvinar receives type II projections (confirmed driver signals) from layer 5 neurons and sends back anatomically defined type I projections (modulatory) to layer 1. In the case of higher-order cortical areas, the assumption is that pulvinar receives type I projections from layer 6 neurons (suspected modulators) and projects to layer 4 (suspected drivers) (Figure 6.1. In the context of predictive coding, this functional distinction between drivers and modulators is crucial, as drivers are associated with predictions and modulators with precision. Precision may coordinate and broadcast more globally visual information and its regulation may be equivalent to selective attention [66], which can be implemented by modulator connectivity from

the pulvinar. This notion of driver and modulators will be discussed further in the first article of this chapter, which reviews additional functional evidences, but also in the second article, which models the heterogeneity of these two types cortico-thalamic pulvinar connections and their role on the propagation of prediction-related alpha oscillations.

6.2. Review: "The Pulvinar as a Hub of Visual Processing and Cortical Integration"

This review for *Trends in Neuroscience* aims to provide an overview of the hypothesized roles of the pulvinar. Given the extensive connectivity, one can find what they would like to find in this nucleus, and thus we aim to review the recent evidences for pulvinar in terms of attention, feature binding, predictive coding, and global workspace theory.

Full citation is as follows: Nelson Cortes, Hugo J Ladret, Reza Abbas-Farishta, et al. "The pulvinar as a hub of visual processing and cortical integration". In: *Trends in Neurosciences* (2023)

As this is a review rather than a research article, we direct the reader to Appendix B. There, they will find a comprehensive overview of the response properties of the pulvinar and their interpretation under predictive processing.

6.3. Methods: Oscillations and Predictions

As we discussed in the introduction and in the review, the pulvinar is involved in performing a gain control modulation onto V1 [252]. This can be conceptualized as a (predictive) gating mechanism, that can control the propagation of message passing from V1. In the case of an overly high variance input, the pulvinar can thus "explain away" the activity of V1, preventing an erroneous update of the internal brain models. But, in the opposite case, what is the influence of V1 on the pulvinar ? If, according to our article in chapter 4, V1 can compute the inverse variance of visual inputs, then it should be able to send that information in some form to the global regulator of the message passing to other cortical areas. Furthermore, if there is such a thing as this global regulator of inverse variance-weighting implemented the pulvinar [153], then it is crucial to look at how V1 it interacts with the pulvinar, in that directed fashion.

To understand the nature of this communication, we must first understand the nature of the message being sent. Disentangling sensory predictions from prediction errors is already arduous enough in V1 and we must turn our attention to another method to study inter-area communications. For that, it is better to study the propagation of activity from the cortex to the pulvinar in terms of frequency. This relies on the idea that, in predictive coding, one posits the existence of a single prediction error unit (not necessarily a neuron) for every prediction unit. As these units must be connected only to one another, and in the form of a negative feedback loop (see

Equation 2.28, predictive coding essentially implies a neural network that works as coupled oscillatory pairs of prediction errors and prediction units [100]. The lowest characteristic frequency of this response, as determined empirically, lies in the alpha range (8 − 12Hz), which is the dominant brain rhythm at rest, when prior models require no novel error updating. This also translates into many psychophysical-relevant peaks in response to a stimulus, as recorded in EEG (N1, P2, for example [40]).
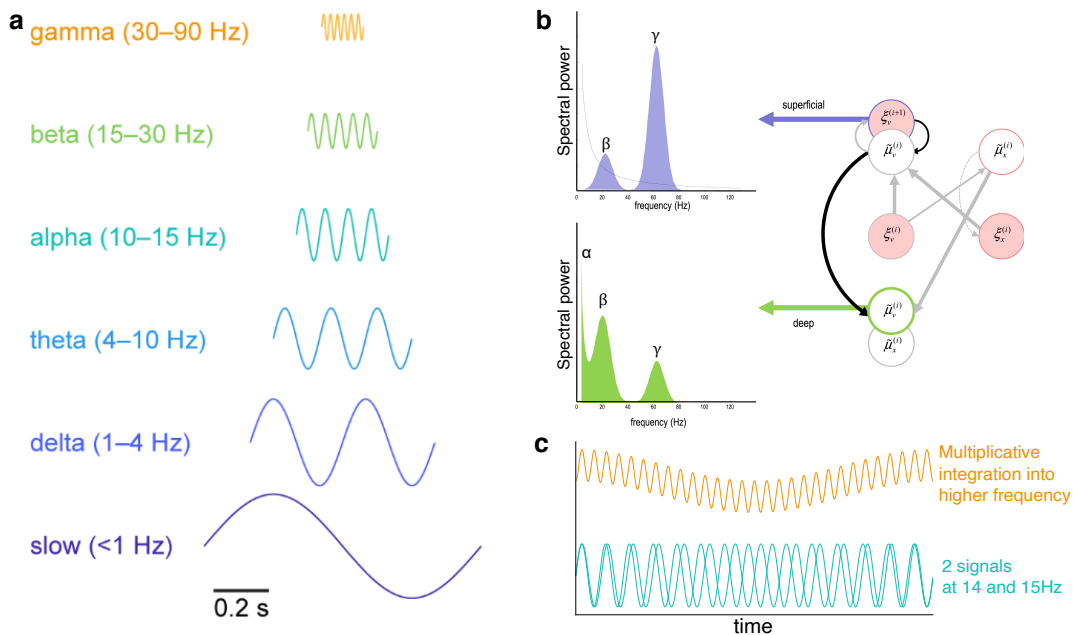
Figure 6.2. – Neurobiological oscillations and predictive coding. (a) Frequencies of oscillations found in neurobiological recordings. (b) Types of oscillations posited by a predictive coding neural network, from [20]. (c) Illustration of the multiplication of two alpha-band predictive signals into a gamma-band prediction error.

The picture becomes more complex when one needs to understand that resolution of erroneous predictions through propagation of errors are not instantaneous. Predictions, fundamentally, should be based on the multimodal interactions of cortical areas, but also of multi-features interactions within a single modality. Practically, a prediction about a high-variance Motion Clouds is the interaction between multiple prediction error units, each signaling a single edge.

This means that the message passing in predictive coding occurs through a series of nonlinear transformation, implying that many alpha range signals will be transformed into higher frequency through this exchange. This summed interaction is often assigned to the second most prominent range of communication in the brain, gamma-band ($40Hz$). Prediction errors, as they are based on this non-linear mixtures of expectations, must be present in those higher frequencies. Predictions, as they are more stable and need not be updated constantly (which forms the defining principle of predictive coding), should be only found in lower frequency. This "spectral

asymmetry" of cortical communication [20, 22] explains many attentional mechanisms [170]. This is also replicated with minimal assumption in silico, by showing that coupled neuronal oscillators with biological time constants, i.e. conduction delays of 12 ms and synaptic time constant of 20 ms naturally create an emergent alpha/gamma prediction/error oscillation.

It follows that spectral asymmetry also implies (functional) anatomical asymmetry, namely in the canonical model that predictions can be found in deep layers [20]. Thus, given the anatomy of the corticothalamic projections, it would seem that predictions only are sent to the pulvinar. This effectively allows it to regulate which prediction of a given level of the visual hierarchy is best suited to provide the best explanation for an external sensory cause, based on its precision. This is specifically the theoretical principle we discuss in this article, by showing how alpha (predictive) rhythms in the pulvinar are gated by cortical activity.

6.4. Article: "Corticothalamic Projections Gate Alpha Rhythms in the Pulvinar"

This theoretical article discusses the role of synaptic asymmetry from the cortex to the pulvinar, so called corticothalamic terminals, and their role in communication between cortical areas. Two types of these terminals, originating from different hierarchical levels of the visual hierarchy, exhibit unique anatomical and functional patterns, causing distinct oscillatory rhythms in the pulvinar. Through a modeled cortical feedforward network including areas 17 (V1) and 21a (V4), we found that these terminal types play antagonistic roles in regulating oscillatory activities in the pulvinar. We suggested that the varying activation of these terminals can gate the pulvinar responses, influencing the oscillatory transfer between lower and higher-order areas, ultimately impacting the neuronal communication throughout the cortical hierarchy. While not emphasized in the article to keep a coherent narrative, this can be interpreted as having a centralized nucleus of the visual network, the pulvinar, ascribing inverse variance weighting to cortical activity in order to modulate the message-passing between alpha (predictions) and gamma (predictions errors) activities.

Full citation is as follows: Nelson Cortes, Reza Abbas Farishta, Hugo J Ladret, et al. "Corticothalamic Projections Gate Alpha Rhythms in the Pulvinar". In: *Frontiers in Cellular Neuroscience* 15 (2021), p. 787170

Corticothalamic Projections Gate Alpha Rhythms in the Pulvinar

Nelson Cortes[1]*, Reza Abbas Farishta[1], Hugo J. Ladret[1,2] and Christian Casanova[1]

[1] Laboratoire des Neurosciences de la Vision, École d'optométrie, Université de Montréal, Montreal, QC, Canada, [2] Institut de Neurosciences de la Timone, UMR 7289, CNRS and Aix-Marseille Université, Marseille, France

Two types of corticothalamic (CT) terminals reach the pulvinar nucleus of the thalamus, and their distribution varies according to the hierarchical level of the cortical area they originate from. While type 2 terminals are more abundant at lower hierarchical levels, terminals from higher cortical areas mostly exhibit type 1 axons. Such terminals also evoke different excitatory postsynaptic potential dynamic profiles, presenting facilitation for type 1 and depression for type 2. As the pulvinar is involved in the oscillatory regulation between intercortical areas, fundamental questions about the role of these different terminal types in the neuronal communication throughout the cortical hierarchy are yielded. Our theoretical results support that the co-action of the two types of terminals produces different oscillatory rhythms in pulvinar neurons. More precisely, terminal types 1 and 2 produce alpha-band oscillations at a specific range of connectivity weights. Such oscillatory activity is generated by an unstable transition of the balanced state network's properties that it is found between the quiescent state and the stable asynchronous spike response state. While CT projections from areas 17 and 21a are arranged in the model as the empirical proportion of terminal types 1 and 2, the actions of these two cortical connections are antagonistic. As area 17 generates low-band oscillatory activity, cortical area 21a shifts pulvinar responses to stable asynchronous spiking activity and vice versa when area 17 produces an asynchronous state. To further investigate such oscillatory effects through corticothalamo-cortical projections, the transthalamic pathway, we created a cortical feedforward network of two cortical areas, 17 and 21a, with CT connections to a pulvinar-like network with two cortico-recipient compartments. With this model, the transthalamic pathway propagates alpha waves from the pulvinar to area 21a. This oscillatory transfer ceases when reciprocal connections from area 21a reach the pulvinar, closing the CT loop. Taken together, results of our model suggest that the pulvinar shows a bi-stable spiking activity, oscillatory or regular asynchronous spiking, whose responses are gated by the different activation of cortico-pulvinar projections from lower to higher-order areas such as areas 17 and 21a.

Keywords: pulvinar, corticothalamic projections, alpha rhythm, cortical oscillations, cortical and subcortical loops, mathematical modeling

INTRODUCTION

Integrating different visual attributes of an image into a single neuronal representation is a difficult task. Throughout evolution, the mammalian visual cortex has solved this computational problem by separating these different features into distinct and parallel processing modules (Bullier, 2001). Interactions between these modules are hierarchical; as more complex levels of organization are created from lower ones (Felleman and Van Essen, 1991; Bullier, 2003; Hegde and Felleman, 2007). This feedforward pathway is accompanied by feedback projections that shape responses to those upcoming signals. Thus, visual processing consists of cortical signals traveling from lower to higher order (HO) areas and vice versa throughout cortico-cortical connections, whose organization follows a hierarchical pattern. This anatomical and functional arrangement of cortical visual areas connected through specific laminar projections is the core of the cortico-centric view of visual integration (Nassi and Callaway, 2009).

Besides direct communication of cortico-cortical areas *via* feedforward and feedback connections, indirect communication through cortico-thalamo-cortical connections also occurs. These transthalamic pathways enable communication between all cortical areas through a limited number of synapses (putatively, only one) in higher order (HO) thalamic nuclei (Sherman and Guillery, 2002; Casanova, 2003). In the visual system of large mammals, the pulvinar is the most prominent HO nuclei and it establishes reciprocal connections with virtually all visual cortical areas of the neocortex (Shipp, 2003; Sherman and Guillery, 2011). This connectivity is reflected in the response properties of pulvinar neurons, which resembles those found in visual cortical cells at different hierarchical levels (Bender, 1983; Casanova, 2003, 2021; Le et al., 2019). It has been recently suggested that the unique network created between the pulvinar and the cortex is used to mediate the temporality of cortical communication (Saalmann et al., 2012; Fiebelkorn et al., 2019; Cortes et al., 2020). For instance, the pulvinar may regulate cortical responses in feedforward and feedback directions by synchronizing distant oscillatory cortical regions, given its strategic position within the visual hierarchy. Such temporal control may be crucial for shaping whole-brain dynamics on a moment-to-moment basis when, for example, attention demand or visual contrast modulation is required (Shipp, 2004; Snow et al., 2009; Cortes and van Vreeswijk, 2012; Cortes et al., 2020; de Souza et al., 2020).

Although the transthalamic pathway is now considered as an essential part of the visual system, how visual processing along this indirect cortical route differs from that along the cortico-cortical pathways has remained elusive (Casanova, 2003). Efforts have been made to determine the anatomo-functional characteristics of the projections between the cortex and the pulvinar. Two types of corticothalamic (CT) projections have been recognized in thalamic nuclei, including the pulvinar. Type 1 axons are thin and have long, thin branches with small terminal endings and are considered to be equivalent to round small (RS) presynaptic terminals observed at ultrastructural level (Rockland, 2019); type 2 axons have thicker axon diameters with clustered endings considered to be equivalent to round

FIGURE 1 | Hierarchical organization and attributes of CT projections. **(A)** Two types of CT projections from lower- mid- and higher-ranked cortical areas target the pulvinar. Note the variation of proportions of type 1 (from layer VI) and 2 (from layer V) axons along the visual cortical hierarchy (**Table 1**). **(B)** Anatomical (left panel) and functional (right panel) classification criteria for projection types 1 and 2. Type 1 terminals have thinner axons, bear small boutons, and elicit EPSP with short-term facilitation; type II projections have thicker axons, larger boutons that can form rosette-like structures and produce EPSP with short-term depression.

large (RL) presynaptic terminals. These terminals originate in different layers: while type 1 projections arise from layer 6, type 2 projections from layer 5 (Vidnyánszky et al., 1994; Ojima et al., 1996; Feig and Harting, 1998; Huppe-Gourgues et al., 2006, 2019). In addition, based on the characterization of their excitatory postsynaptic potentials (EPSPs), type 1 and 2 CT projections display frequency-dependent facilitation and depression, respectively (Li et al., 2003). While type 1 terminals are associated with CT projections from area 17 to LGN, type 2 terminals are more abundant in CT projections from area 17 to the pulvinar (Abbas-Farishta et al., 2020). Altogether, these findings suggest that the types 1 and 2 associated pathways appear to complement synergistically each other to fine-tune visual processing passing by the pulvinar (**Figure 1**).

Although the pulvinar receives more type 2 than type 1 axon terminals from area 17, this ratio is not fixed within the visual cortex. Type 1 endings seem to be more represented in CT connections from higher hierarchical levels (Abbas-Farishta et al., 2020). For instance, in cats, CT terminals emerging from HO cortical areas, as areas 21a [considered to be a homolog of primate area V4 (Payne, 1993)] and the posteromedial lateral cortex (PMLS, the homolog of area MT in primates (Payne, 1993; Huppe-Gourgues et al., 2019), display more type 1 terminals. Furthermore, in the anterior ectosylvian visual area (AEV), one of the highest areas in the hierarchical organization of the visual system, the proportion of type 1 endings highly

6. Beyond V1: Variance and Thalamo-Cortical Loops – 6.4. Article: "Corticothalamic Projections Gate Alpha Rhythms in the Pulvinar"

TABLE 1 | Percentage of CT terminal types as a function of cortical source for areas 17, PMLS, 21a and AEV in cats.

Area	Type 1 (%)	Type 2 (%)
17	25	75
21a	81	19
PMLS	71	29
AEV*	91	9

*Terminals located in the pulvinar subdivision Lateral Posterior lateral (LPl), except for * which comes from Lateral Posterior medial (LPm) (Abbas-Farishta et al., 2020).*

dominates the CT pathway toward the pulvinar (**Table 1**). These findings indicate that the ratio of type 1/type 2 cortico-pulvinar projections increases as a function of the hierarchical position of the source cortical area.

This organizational scheme of CT terminals raises questions about their function and, consequently, the role that pulvinar might play in transthalamic cortical communication. On the one hand, theoretical works have shown that a population of neurons receiving type 1 and 2 terminals exhibit rhythmic or regular firing rate responses given the short-term plasticity dynamics that those terminals have (Tsodyks et al., 1998). On the other, as experimental data shows, irregular spiking activity (Chalupa, 1991; Yu et al., 2018; Wilke et al., 2009) and low-oscillatory rhythms (Saalmann et al., 2012; Fiebelkorn et al., 2019; Halgren et al., 2019) have been detected in the pulvinar. Within the slow rhythms, the alpha-band oscillations (7.5–12.5 Hz) associated with thalamic activity are particularly interesting, as animals performing visual attentional tasks show that the pulvinar drives cortical alpha rhythms (Saalmann et al., 2012), and that cortical alpha waves amplitude decreases when the pulvinar is inhibited (Zhou et al., 2016). One might ask whether the hierarchical gradient of CT terminal types 1 and 2 toward the pulvinar is responsible for generating rhythmic or irregular spiking responses (asynchronous state) in the thalamus. Therefore, to investigate whether CT projection types along the visual hierarchy influence pulvinar neuronal temporal responses differently, we simulated a pulvinar-like network of excitatory and inhibitory neurons receiving terminal types 1 and 2 from two cortical structures simulating a low and a higher-ranked cortical area. The distribution of these corticopulvinar terminals was established using projection patterns of area 17 and extra-striate area 21a of cats, whose anatomy and functional connectivity with the pulvinar have been well documented (Abbas-Farishta et al., 2020, 2021; Cortes et al., 2020; de Souza et al., 2020, 2021). Corticopulvinar projections were implemented with short-term plasticity dynamics, in which terminal types 1 and 2 had facilitation and depression of their EPSP, respectively (Li et al., 2003). Thus, connections were established to reproduce alpha-band oscillations in pulvinar neurons' populations. We found that alpha rhythms in the pulvinar are generated by each cortical area separately or by the simultaneous combination of the two areas in a specific range of connectivity weights. In the first case, when each cortical area independently evokes pulvinar alpha waves, the oscillatory low-frequency activity generated by one area was changed by asynchronous spiking

activity when the other targets the pulvinar. This property suggests that the pulvinar has a bi-stable state of oscillatory or asynchronous responses depending on the origin of the activation coming from CT afferent projections along the visual cortical hierarchy.

MATERIALS AND METHODS

Network Models

Three models were used in this work to simulate the neuronal pulvinar dynamics evoked by cortico-pulvinar projections (**Figure 2A**). Each model is an upgrade of the previous one to provide an integrative framework in which the role of CT projections was investigated. The first model consists of a network in the balanced excitatory-inhibitory (*E-I*) state, in which inputs are modeled as Poisson spike trains. Such inputs have types 1 and 2 dynamics, inducing Short-Term Facilitation (STF) and Short-Term Depression (STD) responses, respectively (**Figures 2B,C**) (Li et al., 2013). With this model, the proportion of input synapses and their strength were varied to investigate how this network, our "protopulvinar," evoked either oscillatory or asynchronous responses. The second model is similar to the previous. However, here, the "pulvinar" network was targeted by two simulated cortical areas, each of them with a combination of type 1 and 2 synapses. As in the previous model, cells were simulated as Poisson spike trains (**Figure 2A2**). The proportion of type 1 and 2 projections, as well as the synaptic contact of those projections to *E-I* pulvinar neurons, were determined by available empirical data from the anatomy of cortical-pulvinar cat projections (**Tables 1**, **2**). This model was useful to identify activation dynamic ranges of thalamic neuron populations as cortical projections reach them.

Finally, the pulvinar network was targeted by other two independent networks that imitate cortical areas 17 and 21a (**Figure 2A3**). Each cortical area was organized to reach the *E-I* balanced state. A feedforward connection, from areas 17 to 21a, was established. Also, feedforward CT projections from the two cortical areas to the pulvinar-like structure were implemented. These cortical axon terminals were simulated with short-term plasticity dynamics, but synapses from the LGN to the area 17 (modeled as Poisson spike trains) and between areas 17 and 21a had linear integration of their synaptic inputs. Cortico-pulvinar projections with types 1 and 2 terminals were chosen randomly from neurons in areas 17 and 21a, and their proportion and contact to *E-I* pulvinar neurons were organized as in the previous model. This model does not consider direct feedforward connections from LGN to area 21a (Wimborne and Henry, 1992). Each area, including the pulvinar network, was simulated with *N* = 10,000 neurons (Vogels and Abbott, 2009; Cortes and van Vreeswijk, 2015; de Souza et al., 2020). From this number of neurons, 20% were inhibitory cells (Rodney et al., 2004). The effect of separated cortico-recipient zones in the pulvinar and their consequences in the response of 21a neurons were studied with this model.

140

FIGURE 2 | Models and short-term plasticity for synapses used in this work. **(A)** Three models were used during simulations. All models contain a pulvinar structure with 80% of excitatory (red dots) and 20 % of inhibitory (blue dots) neurons as represented in the proportion of dots inside boxes. **(A1)** Pulvinar receives only terminal types 1 and 2. **(A2)** Two areas with proportion and number of contacts of terminal types 1 and 2 in areas 17 and 21a as described in **Tables 1, 2**. **(A3)** Cortico-pulvinar-cortical model consisting of three networks connected feedforwardly by cortico-cortical pathways and with a transthalamic pathway. A similar proportion of connections as in panel **(A2)** was used. **(B)** Short-term plasticity for a pulvinar neuron receiving terminal types 1 **(B1)** and 2 **(B2)**. Four input frequencies are tested (2, 5, 10, 20 Hz) for each terminal type. **(C)** EPSP amplitude (mV) as a function of repetitions for five frequency-dependent stimulation regimes and the two types of cortical terminals shown in **(C1,C2)**. Note that the result of the stimulation at 0.5 Hz (blue line) is shown here but not plotted in panel **(B)**.

Neuron and Synaptic Dynamics

Thalamic and cortical neurons were modeled with adaptive exponential integrate-and-fire dynamics (Brette and Gerstner, 2005). This model consists of two coupled differential equations describing the leak current (linear component) and the spike generation component (exponential function). Also, an adaptive current, w, was added to simulate action potential adaptation. The membrane potential of a neuron (i, A, β) is given by:

$$C_m \frac{dV_i^{A,\beta}}{dt} = -g_L^{A,\beta}\left(V_i^{A,\beta} - V_L\right) + g_L^{A,\beta} \triangle_T \exp\left(\frac{V_i^{A,\beta} - V_T}{\triangle_T}\right)$$
$$-w_i^{A,\beta} + I_{input,i}^{A,\beta}, \tag{1}$$

$$\frac{dw_i^{A,\beta}}{dt} = \frac{a\left(V_i^{A,\beta} - V_L\right) + w_i^{A,\beta}}{\tau_{adapt}},$$

where C_m is the capacitance of the neuron, V_L is the leak reversal potential, V_T is the threshold and $\triangle_T$ is the slope factor, τ_{adapt} is the time constant and a describes the level of subthreshold

TABLE 2 | Percentage of contacts from layers 5 and 6 of areas 17 and PMLS to excitatory and inhibitory cells in the pulvinar (Vidnyánszky et al., 1994).

Layer	Inhibitory (%)	Excitatory (%)	Area
5	44.5	48.0	17
5	45	55	PMLS
6	14	85	PMLS

adaptation. Every time that the neuron i fires, w is increased by a current b (spike-triggered adaptation), and the membrane potential is reset to a fixed voltage, V_r of the neuron, i, which has, A, excitatory or inhibitory actions, and determined as cortical of thalamic component, β. Only excitatory neurons have adaptation current dynamics.

The input current that a neuron (i, A, β) receives is:

$$I_{input,i}^{A,\beta} = I_{rec,i}^{A,\beta} + I_{ext,i}^{A,\beta} \tag{2}$$

where $I_{rec,i}^{A,\beta}$ characterizes the synaptic current from recurrent connections of each area β, of a neuron i, with A, E or I, attributes. When cortical cells are integrated in a network, the external

Beyond V1: Variance and Thalamo-Cortical Loops – 6.4. Article: "Corticothalamic Projections Gate Alpha Rhythms in the Pulvinar"

6. Beyond V1: Variance and Thalamo-Cortical Loops – 6.4. Article: "Corticothalamic Projections Gate Alpha Rhythms in the Pulvinar"
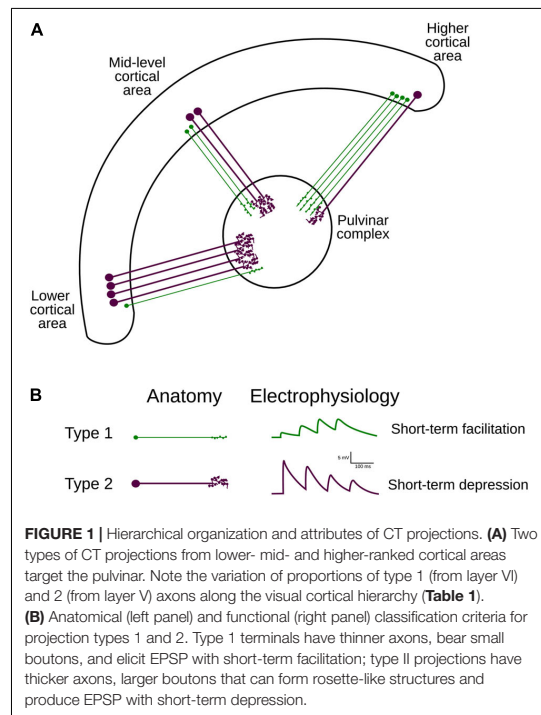
Cortes et al.

CT Projections Gate Pulvinar Oscillations

current, $I^{A,\beta}_{ext,i}$ comprises one term if the unit comes from area 17 and two terms, if it comes from area 21a or pulvinar.

Synapses for cortical and pulvinar recurrent connections and from area 17 to 21a (feedforward pathway) are simulated as an instantaneous rise of synaptic current followed by an exponential decay.

Short-term synaptic plasticity (STP) is implemented with a phenomenological model (Stimberg et al., 2019b). This model considers synaptic release as the product of two variables, x_s and u_s, where x_s represents the fraction of the total neurotransmitter that remains available for release, and u_s reflects the fraction of available resources ready for use, that is, the resources of the neurotransmitter "docked" for release by exocytosis by calcium sensors. After an action potential and the beginning of another, u_s decays to 0 at rate ω_f, and x_s recovers to 1 at rate ω_d, as:

$$\frac{du_s}{dt} = -\omega_f u_s, \tag{3}$$

$$\frac{dx_s}{dt} = \omega_d (1 - x_s). $$

The influx of calcium in the terminal triggered by the arriving of action potentials modifies a fraction U_0 of neurotransmitter resources not expected for release $(1 - u_s)$ to the "docked" state ready to be released (u_S). Eventually, a release r_s from the fraction of u_s of the available neurotransmitter resources are generated, while x_s decreased by the same quantity, so:

$$u_s \leftarrow u_s + U_0 (1 - u_s), $$
$$r_s \leftarrow u_s x_s, \tag{4}$$
$$x_s \leftarrow x_s - r_s. $$

When a presynaptic action potential arrives, A synapses increase the A conductance, g^A_{STP} in the postsynaptic neuron as $g^A_{STP} \leftarrow g^A_{STP} + G_{A0} r_s$, where A = E, I and G_{A0} is the synaptic conductance. Only excitatory connections as short-term plasticity dynamics, given that only excitatory long-range cortico-cortical and cortico-pulvinar terminals have been described.

Feedforward and Recurrent Connections

Recurrent and external connectivity for each structure were random with connection probability, p, specific to E and I populations ($p_{AI} = 0.1$, $p_{AE} = 0.5$). Input current is defined as $I^A_{\beta \leftarrow \gamma}(t) = \bar{g}_{\beta \leftarrow \gamma}(t) \left(V^A_i - V_{\beta \leftarrow \gamma A}\right)$, where the term on the right side of the equation is the sum of all conductance from all presynaptic inputs on the neuron (i, A, γ), where γ is the source and β the targeting structure. In general, it is described as:

$$g^A_{\beta \leftarrow \gamma, i}(t) \quad \frac{\bar{g}^A_{\beta \leftarrow \gamma}}{\tau^A_{syn}} \sum_{j=1}^{N_\gamma} C^{\beta A \leftarrow \gamma A, l}_{ij} \sum_k e^{-\left(t - t^\gamma_{j,k}\right)/\tau^A_{syn}}, \tag{5}$$

where $t^\gamma_{j,k}$ is the time of the kth action potential of the neuron (j, γ). For feedforward external inputs, γ can be the LGN, area 17 or the pulvinar, and β can be the pulvinar, area

17 or area 21a, and $A = E$. For the recurrent connectivity in the pulvinar and the cortex, $\beta = \gamma$, and $i \neq j$. The connection matrices $C^{\beta A \leftarrow \gamma A, l}_{ij}$, for $A = E, I$, are random with probability $c_{\beta \leftarrow \gamma} K / N_{\beta \leftarrow \gamma}$ and $C^{\beta A \leftarrow \gamma E, l}_{ij} = 0$ otherwise. On average, neurons type β receive $K_{\beta \leftarrow \gamma} = c_{\beta \leftarrow \gamma} K$ presynaptic connections from γ neurons. Here, the conductance $\bar{g}_{\beta \leftarrow \gamma}$ describes the strength of the presynaptic input, which is scaled by K as $\bar{g}^A_{\beta \leftarrow \gamma} = G^A_{\beta \leftarrow \gamma} / \sqrt{K}$, where $G^A_{\beta \leftarrow \gamma}$ is independent of K.

For background synaptic activity, a population of simulated neurons ($N_{bckgrnd} = 8,000$, $p_{ANoise} = 0.5$) where $A = E, I$ as a Poisson-type spike train is applied to the pulvinar following the same synaptic dynamics of equation 5. The spike train is excitatory and activates excitatory and inhibitory neurons with a discharge rate of 0.1 sp/s.

When invoking STP for the CT pathway (γ = *areas 17 or 21a*, β = pulvinar), since conductance changes over time, the voltage integration assumes an effective synaptic weight, $g^{A,STP} \beta \leftarrow \gamma$ rather than the static synaptic weights $\bar{g}^A_{\beta \leftarrow \gamma}$.

Parameters

The parameters for the cell dynamics were Cm = 1 μ F/cm², with conductance of leak currents of gL, E = 0.1 mS/cm² and gL, I = 0.05 mS/cm² for excitatory and inhibitory neurons, respectively. The other parameters that characterized the dynamic of neurons with a regular spiking are: VL = −70.6 mV, V_T = −50.4 mV and ΔT = 2 mV. The parameters for the adaptation current were a = 24 nS, b = 0.01 nA, and τ_{adapt} = 60 ms. For bursting $V_L = V_T + 5$ mV, and τ_{adapt} = 20 ms, a = 4 nS, and b = 0.5 nA. For each area, the synapses' parameters were G_{E0} = 1.425 ms nS/cm². G_{I0} = 1.89 ms nS/cm², G_{EI} = 9.0 ms nS/cm², G_{II} = 13.5 ms nS/cm², G_{EE} = 22.5 ms pS/cm², G_{IE} = 67.5 ms pS/cm², with τ_{syn} = 3 ms and V_E = 0 mV and V_I = −80 mV (de Souza et al., 2020). For STP, parameters were settled to obtain similar synaptic performance from experimental data of terminal types 1 and 2 (Li et al., 2013). Therefore, synaptic release probability at rest U_0^{type1} = 0.006 and U_0^{type2} = 0.8; synaptic depression rates ω^{type1}_f = 0.48 s⁻¹ and ω^{type2}_f = 2.0 s⁻¹; synaptic facilitation rate ω^{type1}_d = 1.5 s⁻¹ and ω^{type2}_d = 3.33 s⁻¹; and, the synaptic conductance G_{A0}, for A = E, I. Recurrent connectivity for each area (pulvinar, areas 17 and 21a when are modeled) is K, and the probability of connection was $p_A = K_A / N_A$, for A = E, I.

Variation of Pathway Connections

We used the factors W_{FF} = 5 and W_{CP} = 1.5 to change the weights of feedforward and cortico-pulvinar projections. These factors multiply the ratio G_{E0}/G_{I0} for those entry inputs.

Simulations of network architecture and neuron equations were performed with Python version 3.2 using Brian2 simulator (Stimberg et al., 2019a). Euler's integration was implemented using a time step of 0.05 ms. The accuracy of the results was verified by repeating simulations with smaller time steps (0.025 ms).

Frontiers in Cellular Neuroscience | www.frontiersin.org   **5**   December 2021 | Volume 15 | Article 787170

142

RESULTS

As stated above, three models were used to investigate the oscillatory gating generated by CT terminals in the pulvinar. The first model analyzed the effect of the combination between types 1 and 2 terminals that target excitatory and inhibitory (*E-I*) neurons, whose responses are in the balanced state (**Figure 2A1**). This model provided a basic approximation of the weight ranges of cortico-pulvinar connections that produced oscillatory and asynchronous neuronal responses in this network, our simulated pulvinar. The second model also consisted of external projections to a pulvinar-like structure. However, this model considered two external areas, cortical areas 17 and 21a, whose cortico-pulvinar projections contain a combination of types 1 and 2 synapses (**Figure 2A2**). The fraction of types 1 and 2 terminals and the percentage of synaptic contact reaching *E-I* pulvinar neurons were arranged with available empirical data (**Tables 1, 2**). This model allowed investigating the dynamics of pulvinar responses when the activation between cortical projections from areas 17 and 21a was temporarily deferred. While the first two networks used cortical inputs as Poisson spike trains, the third model simulated explicitly cortical neurons. Here, two similar networks of *E-I* neurons were implemented and connected feedforwardly to reproduce the interaction between areas 17 and 21a (**Figure 2A3**). Each cortical area targets pulvinar neurons, with the proportion and axon terminal contacts settled in the second model. Furthermore, for this model, cortico-pulvinar projections were divided in striate- and extrastriate-recipient zones to investigate the effect of different signaling pattering of the transthalamic pathway across cortical areas (**Figure 2A3**).

Pulvinar EPSPs Evoked by Single CT Activation

Before analyzing the complete pulvinar network, synaptic plasticity of excitatory postsynaptic potentials (EPSPs) was simulated to obtain similar qualitatively experimental magnitudes of those found in neurons of the lateral posterior nucleus (Li et al., 2003), the homologous nucleus to the pulvinar in rats. To that end, a single CT fiber contacting a single neuron with exponential integrate-and-fire dynamics was simulated (Mat and Met, Equation 1). CT fibers were implemented with synaptic plasticity that simulates short-term facilitation (STF, **Figure 2B1**) and short-term depression (STD, **Figure 2B2**). Then, EPSPs from the pulvinar neurons were recorded as the cortical afferent fiber was stimulated with four sets of frequency impulses (0.5, 2, 5, 10, 20 Hz) (**Figures 2B,C**). With this set-up, the strength of the CT fiber was investigated, and experimental amplitude of pulvinar EPSPs were recovered.

Two types of CT fiber responses were generated (**Figure 2B**). These simulations revealed that changes of EPSP amplitudes elicited by the stimulus train at various frequencies follow experimental results of thalamic responses at a given set of synaptic parameters. For simulated type 1 CT fibers, the amplitude of EPSPs enhanced as the stimulation increased in frequency, depicting a saturation in the response after seven consecutive impulses (**Figure 2C1**). A contrary response was

seen in type 2 CT fibers. Here, EPSP amplitudes decreases at higher frequencies, showing a constant response of amplitude after five consecutive impulses (**Figure 2C2**). The amplitude of pulvinar EPSPs showed a similar profile for frequency stimulation higher than 5 Hz. The phenomenon described here is known as short-term plasticity band-pass filtering, in which depression acts as a low-pass filter and facilitation as a high-pass filter (Izhikevich et al., 2003). The combination of the two synaptic components, STD and STF, generates a frequency-specific resonant output. Thus, the combination of type 1 and 2 terminals will produce a resonance with a different band frequency than that observed when each axon terminal activates pulvinar neurons independently. Taken together these results, our simulations evoked type 1 EPSPs with frequency-dependent facilitation (**Figures 2B1,C1**), and frequency-dependent depression for type 2 EPSPs in pulvinar neurons (**Figures 2B2,C2**).

Pulvinar Network Responses Evoked by Type 1 and 2 Terminals

The next step was to study the effect of CT terminals in a population of *E-I* in the balanced state (**Figure 2A1**). The pulvinar was modeled with sparsely connected neurons but strong connections between *E-I* populations (van Vreeswijk and Sompolinsky, 1996). In this network, excitatory frequency-dependent types 1 and 2 terminals identified in the previous section were feedforwardly connected to excitatory and inhibitory pulvinar neurons. The weights of these external "cortical" synapses, G_{E0} and G_{I0}, were invariant in time. The different responses of the pulvinar network were then analyzed when a factor η amplified the excitatory cortical pathway to *E-I* neurons. The logic of the η factor was to conserve the feedforward ratio (G_{E0}/G_{I0}) and only modulate the amplification of the external pathway to pulvinar neurons (Cortes and van Vreeswijk, 2015; de Souza et al., 2020). η_1 and η_2 were defined to amplify types 1 and 2 synaptic terminals, respectively. So, conductances for type 1 terminals, $g^A_{STF} \leftarrow g^A_{STF} + \eta_{1*}\ G_{A0}r_s$, and for type 2, $g^A_{STD} \leftarrow g^A_{STD} + \eta_{2*}\ G_{A0}r_s$, where A are for E or I processes.

Asynchronous and Synchronous Responses

Two clear states the pulvinar network showed when η was changed: a strong irregular (asynchronous state) or a regular (synchronous state) pattern of spiking activity (**Figure 3**). These two activation states were evoked with a Poisson input spike train of 10 sp/s, in a network that had equal synaptic strengths to *E-I* populations of neurons. To characterize further both regimes, peristimulus-time histograms (PSTHs) and the statistics of spike discharges were analyzed. As **Figure 3A** shows, PSTHs for excitatory and inhibitory neurons reflect the global asynchronous (**Figures 3A1,B1**) and synchronous (**Figures 3A2,B2**) states of the pulvinar network. For the asynchronous state, spike statistics showed an exponential average firing rate and a normal distribution of the coefficient of variation (CV) for the inter-spike interval (ISI), signatures of an irregular activation regime (**Figures 3C1,D1**). Such asynchronous activity in recurrent *E-I* populations of neurons was obtained when the number of

143

FIGURE 3 | Two responses of the pulvinar network: (1) asynchronous and (2) synchronous states. **(A)** PSTH (sp/bin) of excitatory (red line) and inhibitory (blue line) neurons when a spike train with Poisson's statistics of 10 sp/s is applied to the network, **(B)** Raster plot for excitatory (red) and inhibitory (blue) neurons. Only 1,000 cells are shown. **(C)** Distribution of average firing rate (sp/s). Arrows indicate the mean of the distribution. **(D)** Interspike intervals (ISI) of the coefficient of variation (CV). Arrows indicate the mean of the distribution. **(E)** Power spectrum of PSTHs. Arrows show the maximum frequency amplitude. Yellow zones characterize alpha-band oscillation range.

excitatory inputs needed to induce firing was only proportional to $\sqrt{K}$ (van Vreeswijk and Sompolinsky, 1996; Renart et al., 2010; Cortes and van Vreeswijk, 2015). This was not the case for the synchronous state (**Figures 3C2,D2**), in which a predominant oscillatory regime in the PSTH was revealed (**Figure 3B2**). Here,

both the average firing rate and the ISI CV had a narrow distribution profile. These differences between the two pulvinar response states were also observed when the PSTH spectrum of frequencies was measured. In the asynchronous state, a peak of oscillatory activity was observed at low frequencies (<5 Hz),

144

whereas, in the synchronous regime, the oscillatory activity showed a clear peak at 12 Hz, which was within the range of alpha-band oscillatory frequency. The amplitude for this peak was higher for the synchronous than the asynchronous state. In summary, the two states can be visually identified by statistical components of the firing rate and the corresponding PSTH. Thus, an asynchronous state was characterized by a high CV (close to 1) with a low amplitude frequency peak of the PSTH, while the synchronous regime had low average firing rate and CV values, and a high oscillatory frequency amplitude of its PSTH.

To identify in which set of values such neural states occurred, inputs and synaptic feedforward strengths to the pulvinar network were gradually increased. The variations were tested in two networks that had similar strengths of *E-I* connectivity. **Figure 4** shows the result of such simulations. For type 1 terminals, the gradual increase of the input produced a gradual increase of the discharge of the neurons, with a clear asynchronous state of the network (**Figures 4A1,B1**). This gradual increase also occurred when η_1 increased, in which the average firing rate and CV increased further (**Figures 4C1,D1**). Another scenario was observed for type 2 connections. Here, the pulvinar network showed a synchronous transition between the inactive and asynchronous states ($\eta_1 \sim 3$). During this transition, the firing rate as well as the oscillation amplitude increased to high value inputs, while their CV was low. Note that this oscillatory transition also occurred for type 1 synapses, but this was less pronounced. Since type 2 terminals have a first EPSP with a larger synaptic response (**Figures 2B,C**), these terminals evoked the oscillatory transition at lower connection intensities than those simulated for type 1 connections.

Oscillatory Responses Evoked in the Pulvinar by Lower and Higher Cortical Areas

While the distribution of cortico-pulvinar terminal types 1 and 2 seems to be hierarchical-level-dependent, the proportion of contacts to excitatory and inhibitory neurons seem to be terminal-type-dependent. For area 17, the distribution of terminal types 1 and 2 are 25–75% respectively. In higher cortical levels, such as area 21a, this distribution is almost reversed in which types 1 and 2 are 81 and 19% of the cortico-pulvinar connections, respectively (**Table 1**; Abbas-Farishta et al., 2020). On the other hand, cortico-pulvinar type 1 terminals contact 85% and 14% of excitatory and inhibitory cells, respectively, type 2 synapses 48 and 44.5%, respectively (**Table 2**; Vidnyánszky et al., 1994). In fact, round large (RL) terminals from area 17 to the pulvinar are predominantly located in the striate recipient zone. For extrastriate cortical areas, terminal zones are characterized as small boutons (RS) (Huppe-Gourgues et al., 2006). Thus, CT projections seem to exert different synaptic actions depending on their origin along the cortical visual hierarchy and the type of terminals contacting pulvinar neurons.

Such anatomical attributes were implemented in the following simulations of the pulvinar network. For area 17, contributions from a Poisson spike train were divided in terminal types 1 and 2, with the number of implicit excitatory cells considered

in the feedforward CT pathway being 25% and 75% of *K*, respectively, where *K* is the average number of total projections. For area 21a, the percentage for types 1 and 2 terminals were 81% and 29% of *K*, respectively. Regardless of their cortical area of origin, type 1 terminals had a connection probability with excitatory and inhibitory pulvinar neurons of $p = 0.85$ and $p = 0.14$, respectively, and type 2 terminals probability contacts were $p = 0.48$ and $p = 0.445$, respectively. To that network, η_1 and η_2 were increased gradually and input of 10 sp/s was applied to analyze the performance of the pulvinar network. The representative frequency, amplitude of this frequency, firing rate and CV were collected after 1 s of simulation, as shown in **Figure 5**.

Effect of Single Cortical Activation

Regardless of whether the input came from area 17 or 21a, at a given strength of the feedforward cortical pathway, the pulvinar network evoked oscillations in the frequency range of 7.5–12.5 Hz (alpha waves). These oscillations were generated at the transition between the quiescent and the asynchronous steady state of the network. In the transition, the network showed a maximum frequency of $\sim$25 Hz (beta- waves). Alpha frequency bands were found besides such maximum frequency oscillation (**Figure 5**, red lines). Alpha waves were symmetrical at the border of the beta bands, but for η_2 high and η_1 low, only one side of the transition showed alpha band responses. For areas 17 and 21a, alpha waves were around $\eta_1 = 13$ and $\eta_2 = 3$, and $\eta_1 = 7$ and $\eta_2 = 5$, respectively. Between these strengths, decreasing and increasing magnitudes of η_1 and η_2, respectively, allowed a continuous transition where alpha waves were always presented. Note that in order to reach such transition threshold, η_1 was higher in area 17 than in area 21a. The inverse happened for η_2, whose magnitude was lower in area 17 than in area 21a. Such transition threshold indicates that the proportion and the distribution of synaptic contacts of terminal types 1 and 2 settled for areas 17 and 21a promote lower values of the cortico-pulvinar connection to reach oscillatory alpha-band activity.

Effect on Alpha Waves of Sequential Activation of Two Cortical Areas

Here, alpha-band oscillations generated by the driven area were measured when the other area was activated afterward. To elicit alpha rhythms that were representative of cortical areas 17 and 21a, we selected strengths of their connections such that $\eta_2 \geq \eta_1$, and $\eta_1 > \eta_2$, respectively. In this setting, the thalamic network was simulated for 5 s. After the network reached a stable alpha-band oscillation induced by the driven area (2 s), the other cortical was "attached" (1 s). Subsequently, the attached area was disconnected, and a recovery period was allowed (2 s). The results of such simulations are shown in **Figure 6**. When driven inputs were from area 17, alpha waves rose quickly ($\sim$150 ms) inside the network, having a stable oscillatory profile before the end of the first second. The amplitudes of such oscillations were low ($\sim$15 sp/bin), even if pulvinar neurons were synchronized. In the attached period, inputs from area 21a abolished the synchronization, including alpha rhythms. Once 21a was disconnected, the pulvinar network

145

FIGURE 4 | Quantitative outputs for a pulvinar network whose neurons receive terminal types 1 (1) and 2 (2) as input, and connection weights increase gradually. Factors η_1 and η_2 amplify corticothalamic projections of terminal types 1 and 2. **(A)** Frequency (Hz), **(B)** Amplitude (V^2/Hz), **(C)** Firing rate (sp/s), **(D)** CV.

FIGURE 5 | Quantitative outputs for a pulvinar network whose neurons receive projections from areas 17 (1) and 21a (2). Factors η_1 and η_2 amplify corticothalamic projections of terminal types 1 and 2, respectively. **(A)** Distribution and number of contacts to excitatory and inhibitory pulvinar neurons of terminal types 1 and 2 for cortical areas. Red and blue letter colors indicate excitatory and inhibitory percentages, respectively. Terminal types 1 and 2 are represented by colors green and purple, respectively. **(B)** Frequency (Hz), **(C)** Amplitude (V^2/Hz), **(D)** Firing rate (sp/s), **(E)** CV. Red lines in panels **(B–E)** indicate zones where alpha rhythms are evoked.

146

FIGURE 6 | Two solutions to generate alpha waves in the pulvinar. These solutions are found when $\eta_1 = 0.5$ and $\eta_2 = 2.5$ for area 17, and $\eta_1 = 6.0$ and $\eta_2 = 1.0$ for area 21a. Note that the thalamic alpha rhythms are ceased when the other area functionally targets the pulvinar (2–3 s). Solution for areas **(A)** 17 and **(B)** 21a. Terminal types 1 and 2 are represented by colors green and purple, respectively.

came back rapidly to alpha-band oscillations again. When the driven input was from area 21a, the rise of alpha waves was much slower (~1,000 ms), but amplitudes of the oscillation were much larger (~100 sp/bin). Since type 1 synapses have low-amplitude EPSPs in their first pulses, the network takes longer to balance their inputs than when the inputs have EPSPs of high amplitude (Cortes and van Vreeswijk, 2015), as is the case with type 2 connections. Such neuronal states may be similar to those conditions found in the loss of consciousness due to analgesics (Bastos et al., 2021). Adding projections from area 17 induced an asynchronous stable activity on the thalamic population, which returned quickly to alpha waves when this attached area was disconnected. In summary, the pulvinar network developed alpha rhythms by driving cortical inputs and, at these parameter values, the arrival of other cortical sources

generated a global asynchronous state and a loss of oscillatory alpha-band activity.

Figure 6 showed that terminal types 1 and 2 seem to produce different activation dynamics in the thalamic network, particularly at the beginning of the stimulation. As previous pulvinar responses were only analyzed after 1 s of "recording," the aim of the following section was to quantify pulvinar dynamics just before the onset of cortical stimulation. For this purpose, η_1 and η_2 were gradually increased for the strength of projections from areas 17 and 21a. The result of such iterations revealed that pulvinar alpha waves were located in similar activation zones shown above (**Figure 7**). In this regime, pulvinar alpha rhythms appeared and stabilized rapidly when type 2 terminals dominated cortical projections of area 17 (**Figures 7A1,A2**). Conversely, adding type 1 terminals and decreasing the strength

147

FIGURE 7 | Terminal types 1 and 2 produce different activation dynamics in the thalamic network at the beginning of the stimulation. **(A)** PSTH and raster plots for area 17 solutions. **(B)** Outcomes of such solutions for (1) coefficient of variation of alpha waves peaks, (2) number of peaks per second, and (3) first peak onset. **(C)** PSTH and raster plots for area 21a solutions. **(D)** Outcomes of such solutions for (1) coefficient of variation of alpha waves peaks, (2) number of peaks per second, and (3) first peak onset.
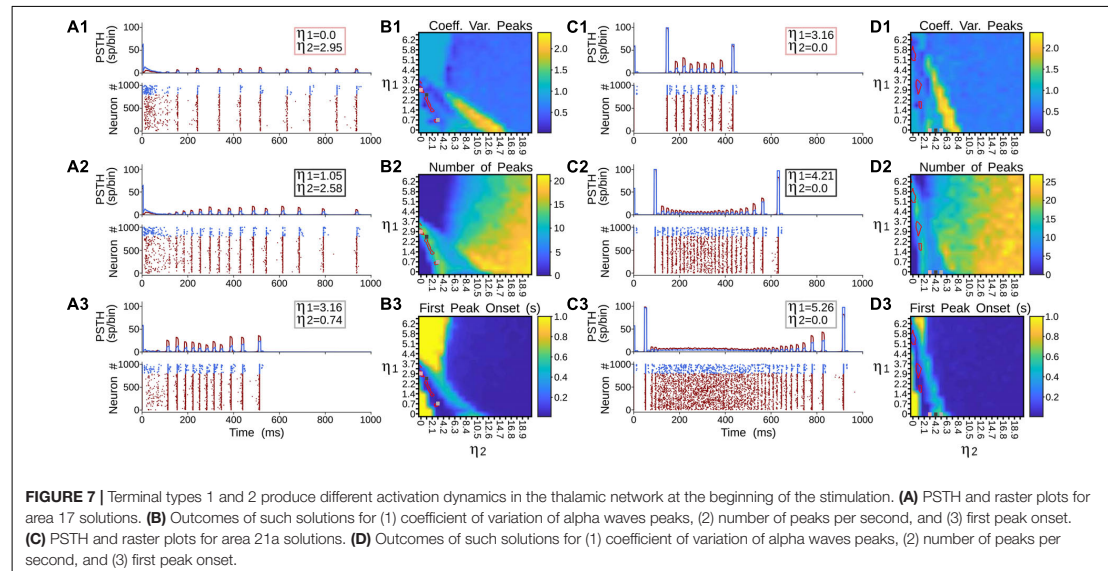
of type 2 axons restricted such oscillations to a short time window (**Figure 7A3**). In fact, increasing the pathway strength of only type 1 axons into the pulvinar network, generated an asynchronous transition of spike activity which was subsequently transformed into a synchronous oscillation when η_1 was large (**Figures 7C1,C3**). These qualitative details were further analyzed by measuring average dispersion (CV), number and the first-time onset of the PSTH peaks of the oscillatory pulvinar synchrony (**Figures 7B–D**). For area 17, alpha-waves were highly regular (low CV) when $\eta_2 > \eta_1$ (**Figure 7B1**), with a constant number of peaks and rapid triggering of activity in similar regimes where alpha rhythms were stable over longer simulation times (**Figures 7B2,B3**). Alpha rhythms for area 21a were also expressed in the same regime (i.e., $\eta_2 > \eta_1$). Here, however, when $\eta_2 < \eta_1$, alpha waves were less evident and stable oscillations with regular periodicity in the first second of simulation were undetected. Thus, in the first period of cortical stimulation, type 2 terminals were more likely to establish effectively a stable and fast alpha-band periodicity than type 1 axons regardless of their source area.

Simultaneous Cortical Activations

The generation of alpha waves in the pulvinar was studied when the two CT projections were combined. The objective here was to evoke alpha waves in the pulvinar with the two cortical areas activated simultaneously. This joint activation may be possible because, for instances, the superposition of oscillatory outputs from areas 17 and 21a (matrices from **Figures 3B1,B2**) generates spots of alpha-wave responses. To that end, the connectivity weights were iterated to find representative examples of cortico-pulvinar projections that achieved alpha rhythms in the thalamus. Such results are shown in **Figure 8**, in which the driver activity

to pulvinar neurons was originated from the simultaneous convergence of the two cortical sources. In **Figure 8A**, the two cortical areas had weights of type 2 connections weaker than those from type 1, whereas, in the **Figure 8B**, terminal types 1 and 2 from area17 were stronger than those from area 21a. The convergence of the two cortical inputs yielded a stable oscillatory alpha-band response after 1 s, which ended when such projections were disconnected (after 2.5 s). To demonstrate that the oscillatory activity was evoked by the synergy of the two cortical areas, the thalamic network was initially connected by only one cortical area. After a period (1.5 s), the input from the other cortical area was restored (**Figures 8A2,A3,B2,B3**). Terminal weights used here were the same as before. In the two above cases, the input from one area was insufficient to gate oscillatory responses in the pulvinar (**Figures 8A2,A3,B2,B3**). When the weight of connections from area 17 to the pulvinar was more robust than those weights from area 21a (**Figure 8B3**), thalamic neurons showed oscillatory responses, but only during a short period (<1.0 s). In this regime of connections, alpha waves in the pulvinar recovered when connections from the disconnected cortical were added from area 17 or area 21a, showing a stable periodicity during the whole period of dual cortical stimulation.

Oscillatory Responses of the Transthalamic Network

In this section, each cortical area was modeled explicitly as a network of *E-I* neurons whose responses were settled in the balanced state (**Figure 2C3**). Randomly chosen excitatory neurons formed the two CT pathways from areas 17 and 21a ($p = 0.2$, homogeneous distribution). The proportion of terminal types 1 and 2 were 25% and 75% and 81% and 29% of

148

FIGURE 8 | Generation of alpha waves by the combination of the two CT projections. **(A)** Solution when type 2 terminals (purple axons) are weaker than type 1 (green axons). **(B)** Solution when terminal types 1 and 2 from area 17 are stronger than connections from area 21a. For panel **(A)**, $\eta^{a17}_1 = \eta^{a21a}_1 = 0.3$ and $\eta^{a17}_2 = \eta^{a21a}_2 = 1.8$. For panel **(B)**, $\eta^{a17}_1 = \eta^{a17}_2 = 2.0$ and $\eta^{a21a}_1 = \eta^{a21a}_2 = 0.5$.
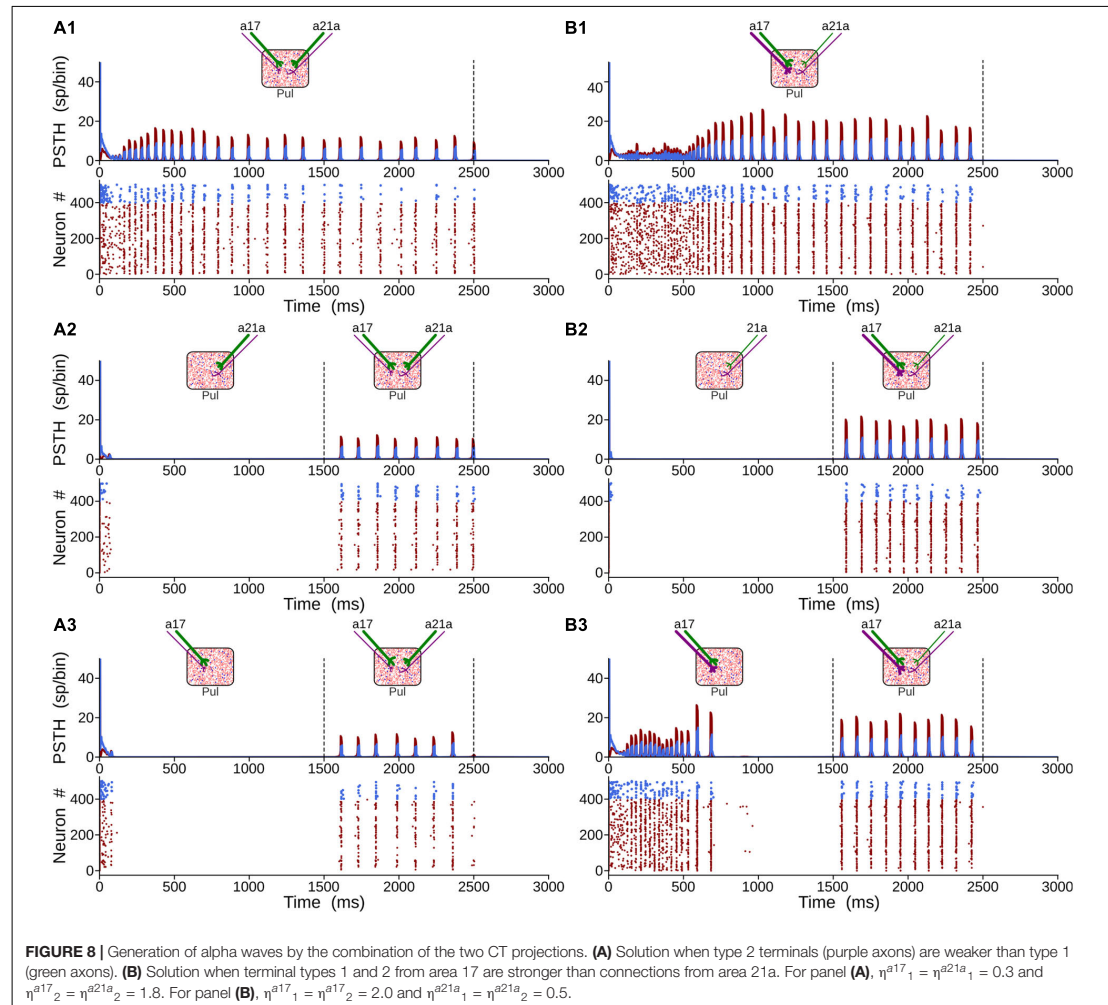
$Np = K$ synapses, for areas 17 and 21a, respectively (**Table 1**). While the proportion of excitatory type 1 terminals contacting excitatory and inhibitory pulvinar neurons was $p = 0.85$ and $p = 0.14$, respectively, for type 2 terminals it was $p = 0.48$ and $p = 0.445$, respectively (**Table 2**). The two cortical areas were also connected feedforwardly through randomly chosen excitatory connections from area 17 to excitatory and inhibitory neurons of area 21a ($p = 0.05$). Inputs to area 17, from an implicit lateral geniculate thalamic nucleus (LGN), were also drawn from a uniform distribution ($p = 0.05$), and were modeled as spike trains with Poisson's statistics. Axon terminals from LGN to area 17, and from area 17 to area 21 did not have any synaptic plasticity. The three interconnected structures mimic the cortico-thalamo-cortical network formed by areas 17 and 21a and the pulvinar.

Effect of Cortical and Pulvinar Dynamics in Oscillatory Responses of the Transthalamic Network

Burst discharges and low background levels of synaptic noise in the pulvinar were first settled to study the cortico-pulvinar pathway of the model. It has been postulated that intrinsic burst neurons in layer 5 (5IB) of the lower cortical areas (i.e., areas 17 and 18) participate in the conduction of pulvinar alpha rhythms (O'Reilly et al., 2021). Furthermore, thalamic bursting has been characterized as an intrinsic property of pulvinar neurons (Ramcharan et al., 2005). Besides, background noise can alter the synaptic efficiency of connections by increasing subthreshold fluctuations (Silver, 2010) and enhancing the salience of neuronal oscillations to avoid long asynchronous/synchronous state transitions as in **Figures 7C1–C3**. For that end, intrinsic bursting was induced in cortical and pulvinar neurons by changing

149

mainly high reset values ($V_r > V_T$), among other parameters of the current regular spike train (Brette and Gerstner, 2005). Furthermore, stochastic balanced synaptic inputs without short-term plasticity were added equally to *E-I* pulvinar neurons to elicit background noise. Here, a simplified version of the neural network described in **Figure 2C3** was used, where only CT projections from area 17 to the pulvinar were incorporated. Thus, the effect of cortical burst spikes and background noise into pulvinar neurons on thalamic alpha waves were compared to cases where no such functional biological processes were present.

Figure 9 shows all cases of combinations where area 17 and pulvinar neurons have regular or burst spiking discharges and pulvinar is with or without background synaptic noise. Raster plots for cortical and pulvinar neurons (top and bottom panels, respectively) and PSTH envelopes for excitatory and inhibitory pulvinar neurons (middle panel) are shown here. To obtain such results, cortical area 17 was targeted by an LGN input of 50 sp/s, and cortical and pulvinar excitatory and inhibitory neurons were recorded for 1 s. Parameters of the brain structures were fixed between simulations, but cortico-pulvinar weights were fitted to obtain pulvinar alpha rhythms. Cases with background noise in the pulvinar were analyzed first, then when neurons from area 17 had bursting discharges, and finally, when thalamic neurons incorporated such burst-like dynamics. Adding background noise to pulvinar excitatory and inhibitory populations of neurons improved the profile of alpha-band oscillations (**Figures 9A,B**). The average dispersion of the excitatory bumps in the PSTH formed around the synchronous discharge of neurons were 30.05 ± 0.14 ms (mean $\pm$ std) and 32.54 ± 0.42 ms (*t*-test, *p-value* = 0.001) for simulations with or without background noise in the pulvinar, respectively. When cortical neurons had bursting dynamics (**Figures 9C,D**), bumps of alpha waves in the pulvinar with and without background, became larger with higher peaks ($\sim$10 sp/bin) and smaller dispersion ($\sim$29.05 ms), compared to previous cases with only regular spiking discharges (Case A vs. Case C, *t*-test *p-value = 0.06*, Case B vs. Case D, *t*-test *p-value = 0.004*). Changing pulvinar regular spiking cells with bursting neurons favored the irregular periodicity of alpha rhythms, but background noise added into excitatory and inhibitory populations stabilized partially such oscillatory lost (**Figures 9E,F**). Taken together, synchronous alpha rhythm responses of pulvinar neurons were affected by cortical and pulvinar neuronal dynamics as well as the background noise added to thalamic neurons.

Pulvinar responses to different cortical and pulvinar dynamics were further investigated by measuring oscillatory frequency and amplitude, as well as firing rates and CV when η_1 and η_2 were iterated (**Figure 10**). Alpha waves were highlighted by depicting contours that had high and low amplitude spectral density in the Fourier transform (**Figure 10**, red and white lines, respectively). In general, alpha waves were between the transition of quiescent states and stable asynchronous activity. While strong alpha-band oscillations were limited to be in such transition (red lines), low energy frequencies were revealed to be more ubiquitous (white lines). Noise decreases the weights used to evoke such strong alpha-frequency oscillatory responses, when comparing iterations with and without background noise (**Figure 10A** vs.

Figure 10B and **Figure 10C** vs. **Figure 10D**). Background noise also restricted the feasible region for finding the alpha waves. On the contrary, incorporating cortical bursting dynamics in the network expanded such regions, particularly for weak alpha-band oscillations (**Figures 10B–E**). Strong alpha rhythms were eliminated when burst responses were included into pulvinar neuron dynamics. As **Figures 9D–E** show, weak alpha waves were located at the transition from silent to activated states, when background synaptic noise was present and cortical and pulvinar neurons had bursting dynamics. Taken together, incorporating bursting cortical and pulvinar dynamics and adding background synaptic noise into pulvinar neurons enhanced oscillatory alpha-band states of the cortico-pulvinar network, but the localization of such low-frequency transition regimes was almost unchanged.

Effect of Transthalamic Pathway in Higher-Order Cortical Responses

Pulvinar effective connectivity to the visual cortex was tested by forming the transthalamic pathway. In this scenario, cortical area 21a received two feedforward projections: one coming from area 17 and another from the pulvinar (**Figure 2C3**). In turn, area 17 received an input from the LGN that consisted of spike trains with Poisson's statistics. Connections from the LGN to area 21a were not implemented. The pulvinar was divided in two functional areas based on the arrangement of CT projections: striate- and extrastriate-recipient zones. Each recipient zone received percentage of connections and excitatory and inhibitory proportion of contacts from terminal types 1 and 2 of cortical areas 17 and 21a as previous simulations. However, connections between *E-I* pulvinar cells were strong, similar to the organization of the two cortical areas. For simplicity, only one pulvino-cortical projection was considered, which was for the connection from the pulvinar to area 21a. For this area, each excitatory and inhibitory populations received in average K random contacts from excitatory pulvinar neurons. Weights of this pulvino-cortical projection, $G^{21a \leftarrow pul}_{E0}$ and $G^{21a \leftarrow pul}_{I0}$, were multiplied by the same factor, W_{CP}^{21a}, which was fixed for all simulations. Dynamics of the pulvinar neurons used here were those established in case A (**Figure 10A**), when the neurons had regular spiking and low background synaptic noise. Thus, the neural network used for the next iterations consisted of a cortico-cortical feedforward pathway from area 17 to area 21a, and a transthalamic pathway from area 17 to the pulvinar, and from pulvinar to area 21a, also defined by CT projections from areas 17 and 21a to striate- and extrastriate-recipient zones in the pulvinar, respectively.

Propagation of Alpha Rhythms From Pulvinar to Area 21a
To achieve alpha waves in the pulvinar, CT projections from area 17 to the pulvinar were fixed to a constant value, and CT weights from area 21a to the pulvinar were gradually varied. **Figure 11** shows an overview of the network performance. For these simulations only, $\eta^{21a}_1 = \eta^{21a}_2$. During the first second, the cortico-cortical pathway was attached, whereas the cortico-thalamo-cortical pathway was unconnected. By consequence, predominant low-frequency oscillations were absent in area 21a. Then, CT from area 17 and pulvino-cortical projections to area

FIGURE 9 | Combinations of cases when area 17 and pulvinar neurons have regular or burst discharges, and pulvinar is with or without background synaptic noise. For each case, raster plots for area 17 and pulvinar and PSTH for pulvinar are shown. Case **(A)** Regular spiking for area 17 and pulvinar neurons, background synaptic noise added to pulvinar neurons. Case **(B)** Similar to case **(A)**, but without background synaptic noise in pulvinar. Case **(C)** Bursting dynamics for area 17, regular spiking for pulvinar neurons, background synaptic noise added to pulvinar neurons. **(D)** Similar to case **(C)**, but pulvinar neurons without background synaptic noise. **(E)** Bursting dynamics for area 17 and pulvinar neurons, background synaptic noise added to pulvinar neurons. **(F)** Similar to case **(E)**, but pulvinar neurons without background synaptic noise. **Supplementary Table 2** shows values of η^{a17}_1 and η^{a17}_2.

21a were added. In this configuration, alpha-band oscillations were generated in cortical area 21a by the transthalamic pathway (1–2 s). Note that during this period, alpha waves were only evoked in the pulvinar striate-recipient zone. In the next temporal sequence (2–6 s), CT connections from area 21a were added into the pulvinar extrastriate-recipient zone, and these projections were amplified gradually throughout this period. Alpha waves in area 21a and the pulvinar persisted when cortical projections

151

FIGURE 10 | Quantitative outputs for cases in **Figure 9**. Frequency (Hz), amplitude (V²/Hz), firing rate (sp/s) and CV for different cases as η_1 and η_2 increase gradually. Red line, alpha waves with higher amplitude; white line, alpha waves with lower amplitude. Refer to the text for more details.
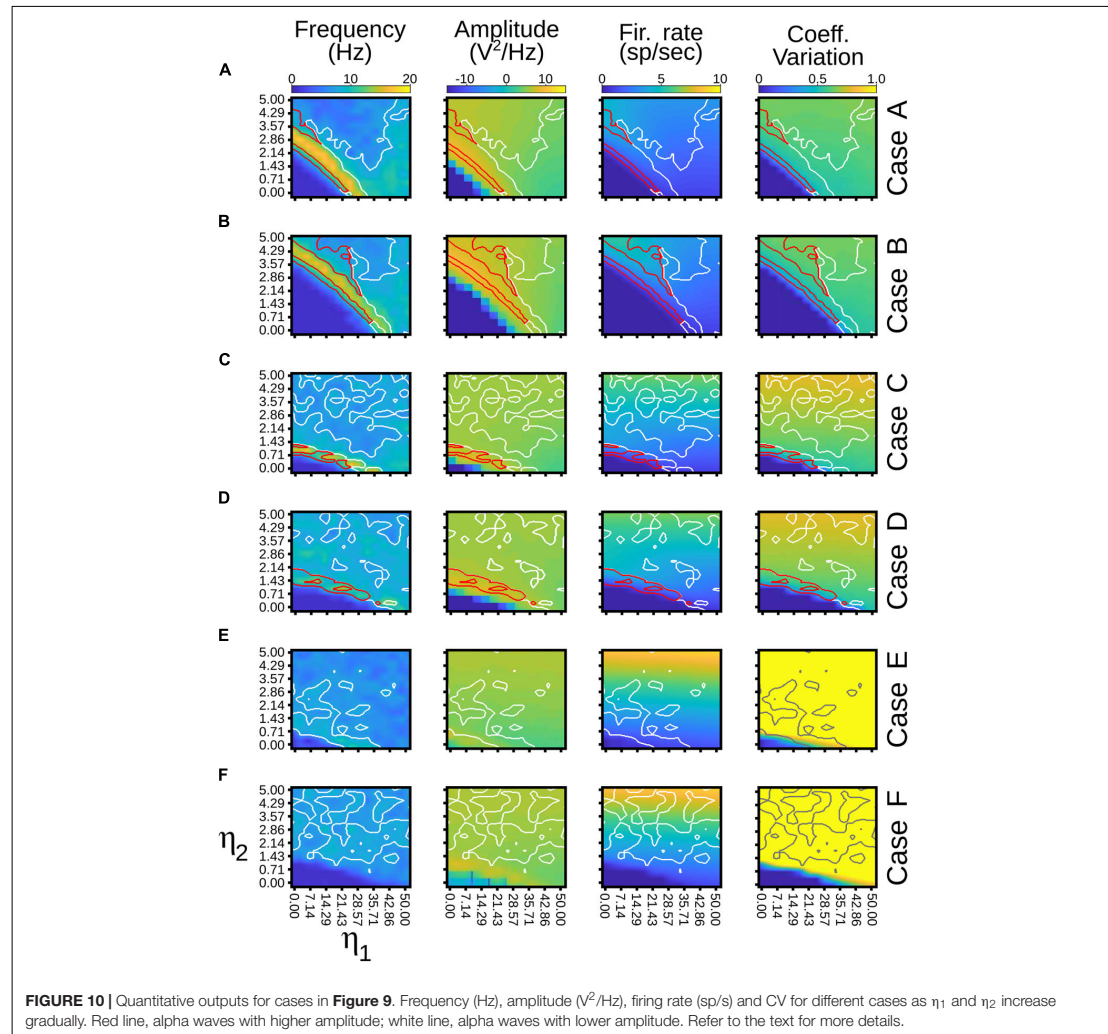
had a low magnitude (2–3 s). In the contrary, alpha-band oscillations were lost when η^{21a}_1 and η^{21a}_2 enhanced (3–6 s). The pulvinar and area 21a showed an asynchronous profile of spike activity when cortical top-down connections to the pulvinar were sufficiently high. Note that pulvinar responses were produced only in the extrastriate-recipient zone, while responses in the striate- recipient zone were silenced. Finally, alpha waves in the pulvinar and area 21a were re-established when cortico-pulvinar connections from area 21a were disconnected (6–7 s). Taken together, the transthalamic pathway enabled functional connectivity in alpha frequency band when cortico-pulvinar connections from lower cortical areas were allowed and an asynchronous response in higher cortical areas when the cortico-pulvinar loop between area 21a and pulvinar was closed.

Closing the Cortico-Pulvinar Cortical Loop

The effect of closing the loop between cortical area 21a and the pulvinar on the formation of alpha waves was further analyzed by varying the weights of terminal types 1 and 2 of these CT projections. For that end, a network incorporating cortico-cortical and transthalamic pathways (area 17 to pulvinar, pulvinar to area 21a) was created as above described. In addition, the network had a top-down projection from area 21a to the pulvinar, which was organized in three different time periods. In the first condition, the projection from area 21a and the pulvinar was disconnected (open loop) for 1 s (**Figure 12A**). The next second, the projection from area 21a to the pulvinar was formed (close loop), and connections from area 17 to the pulvinar were still activated. In the last second, only the close loop was

152

FIGURE 11 | Example of oscillatory alpha-band activity transmission by transthalamic network for a network as **Figure 2A3**. **(A)** Weights of connections, $\eta^{a17,21a}_1$ and $\eta^{a17,21a}_2$ for areas 17 and 21a. PSTH for areas 17 **(B)**, 21a **(C)** and pulvinar **(D)**. **(E)** Raster plots for neurons in the pulvinar. Note that half of the pulvinar neurons are targeted by CT projections from area 17 and the other from area 21a.
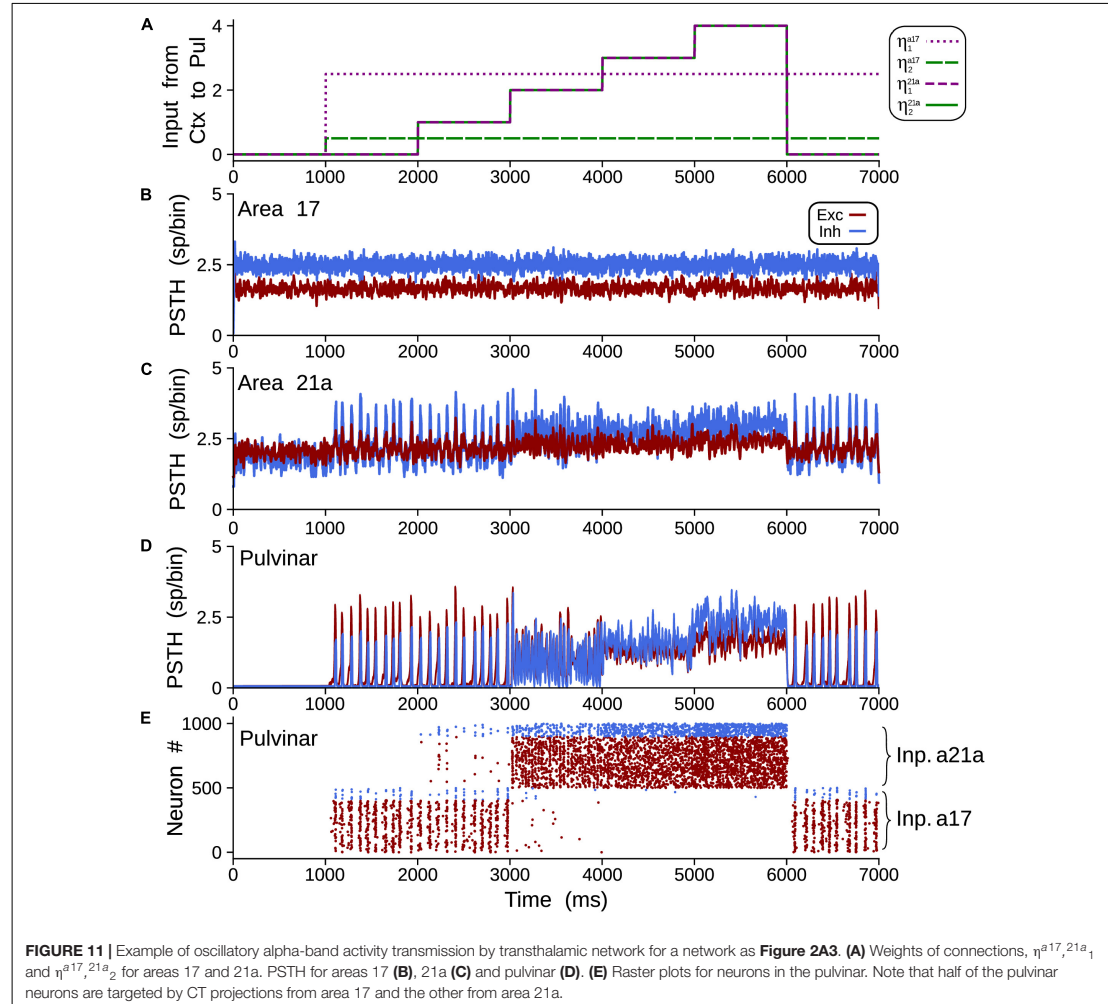
functional, in which cortical top-down activity to the pulvinar was evoked indirectly by cortico-cortical arriving inputs to area 21a (**Figure 12B**).

With this network, the efficacy of alpha waves expressed in area 21a by pulvinar projections was quantified. For the first second, alpha waves in the pulvinar and area 21a were evoked 99% and 64.44% of the time, respectively. The remaining ∼35% of oscillations evoked in area 21a were lower than 7.5 Hz. In the next condition, η^{21a}_1 and η^{21a}_2 were iterated and the frequencies, amplitudes, firing rates and CV of area 21a and pulvinar outputs were characterized. **Figure 12C** shows that, in the second period (2–3 s), closing the loop between area 21a and the pulvinar decreased alpha-band oscillations in the two networks. In fact, the pulvinar achieved only 36% of alpha waves, causing a loss of alpha-wave transmission in area 21a (∼26%).

Alpha-band oscillations in the pulvinar were only expressed at low magnitudes of η^{21a}_1 and η^{21a}_2, since the striate-recipient zone continued to yield such rhythms. When the strength of the cortico-thalamo projections, from area 21a to the pulvinar extrastriate-recipient zone was increased slightly more, the striate-recipient zone was perturbed, and so was the production of alpha waves, decreasing the propagation of such frequency into area 21a during this second period. Note that a variety of higher frequency oscillations in area 21a and the pulvinar arose at such intermediate neuronal states (**Figure 12B1**). As the strengths of the CT connections were higher, only extrastriate-recipient neuronal responses in the pulvinar were engaged, causing an asynchronous spiking state in both the pulvinar and area 21a (**Figure 12B2**). Similar responses were observed in the last period of stimulation (3–4 s). Here, the closed loop

153

FIGURE 12 | Effects on alpha waves when the loop between area 21a and the pulvinar is closed in the cortico-thalamo-cortical model. **(A)** Scheme showing the temporal configuration of the three periods for the connections of the model. Panels **(B,C)** are two solutions when top-down loop to the pulvinar generate asynchronous (1) or oscillatory (2) responses. **(B)** PSTH for area 21a and pulvinar. **(C)** Raster plots for pulvinar neurons with striate- and extrastriate-recipient zones. Second and third periods' quantitative outputs (frequencies, amplitudes, firing rates and CVs) for area 21a **(D)** and **(E)** pulvinar networks when η^{21a}_1 and η^{21a}_2 are increased gradually. Note the position of solutions 1 and 2 in the panel **(D)**.

FIGURE 13 | Schematic representation of the oscillatory responses predicted by our models. **(A)** The pulvinar evokes an alpha-band oscillatory response by the feedforward activity of area 17. The response of the infragranular layers of area 17 is evoked by the inter-laminar connectivity that initially reaches layer 4. **(B)** Oscillatory alpha-band responses in the pulvinar are evoked by feedback activity from higher order (HO) cortical areas to the infragranular layers of area 21a. **(C)** Indirect oscillatory alpha-band activation of the pulvinar by the feedback from area 21a to the infragranular layers of area 17.

was still coupled, but connections from area 17 to the pulvinar were disconnected. Under this configuration, only 1.7% of alpha waves were generated in area 21a. In average, the pulvinar did

not show any alpha waves during this period. The remaining spiking responses of area 21a and the pulvinar had almost all asynchronous features or higher-frequency oscillations. More

154

details of the activity throughout these periods are shown in **Supplementary Table 1**. Altogether, similar to previous analysis, the transthalamic pathway allowed a propagation of alpha waves from the pulvinar to area 21a, which was occluded by the arriving of the top-down projection from area 21a to the pulvinar. Thus, the pulvinar showed a bi-stable spiking response, oscillatory or irregular spiking, whose responses were gated by the different activation of projections from areas 17 and 21a to the pulvinar.

DISCUSSION

Previous experimental studies have shown that neuronal populations of the pulvinar nucleus possess multiple dynamic activity profiles (Wrobel et al., 2007; Saalmann and Kastner, 2011; Yu et al., 2018; Le et al., 2019). In this work, using a theoretical approach, we propose that such population dynamics in the pulvinar may arise from CT connections originating from different hierarchical cortical areas, whose axonal terminals have distinct anatomical and physiological profiles. Two types of neuronal discharges were evoked in the pulvinar, synchronous and asynchronous responses, when projections from areas 17 and 21a to the pulvinar were organized according to the connection empirically described in cats (Huppe-Gourgues et al., 2006, 2019; Abbas-Farishta et al., 2020). We found that, at a specific connection weight regime, CT connections reproduce the oscillatory alpha-band activity of the pulvinar. This solution from the model was found when each cortical area independently contacted the pulvinar and when they were combined. When a single area evoked oscillatory alpha-band activity in the pulvinar, the activation of the distant area ceased such oscillations. The combinations of the two projections types 1 and 2 changed the oscillatory dynamics to asynchronous spiking states. Similar bi-stable state has been reported experimentally in the primate pulvinar during object detection and passive viewing (Wilke et al., 2009). The results found in our models are a direct consequence of the restrictions we put on the parameters, suggesting that the pulvinar may have several dynamic response states to interact with the visual cortex.

The most important prediction of this work is that the pulvinar has at least two functional response states, regular oscillatory or stable asynchronous. Other models explaining such oscillatory variations have included alpha waves in the pulvinar implicitly (Quax et al., 2017) or only as part of the cortico-cortical circuitry (Jaramillo et al., 2019). Our model suggests that part of such oscillatory activity comes from types 1 and 2 CT projections distributed as a gradient along the hierarchy of the visual system. While alpha rhythms in the pulvinar may reflect the functional cortical feedback connectivity (Halgren et al., 2019), our work is in agreement with the experimental evidence that the pulvinar drives cortical alpha waves (Saalmann et al., 2012). In addition, our model predicts that such oscillations may be abolished by CT projections from lower cortical areas (areas 17 and 18). As shown here, given that such connections have a high proportion of type 2 connections, their role would be to desynchronize the pulvinar when activated in a synchronous manner. Although, additional anatomical and physiological data are needed to confirm our

hypothesis, the implications of such predictions are explained in more detail below.

Mechanism of Generating Oscillations

To obtain oscillatory alpha-band activity, in our model, we settled connections to be unbalanced, so CT weights were just lower than $\sqrt{K}$ synapses providing low-frequency waves. Since oscillatory solutions were in the frontier of a balanced state (**Figure 6**), the addition of an extra input generated an asynchronous irregular spiking activity. Such a neural property of networks in the balanced state explains the sharp transition between the two states, which could have implications for fast temporal neural encodings in the pulvinar, as recently proposed (O'Reilly et al., 2021).

Mechanisms that generate alpha rhythms have been attributed to collateral branches of CT axons to the thalamic reticular nucleus (TRN), which in turn provide inhibitory inputs back to the thalamus. Such glutamatergic excitatory input from cortical layer VI targets excitatory and inhibitory TRN cells, while the latter component, GABAergic neurons, hyperpolarized thalamic excitatory relay cells that discharge action potentials in burst firing rate. The burst stimulates collateral TRN cells and serves to re-inhibit excitatory relay cells. This loop continues in a low-frequency oscillation at 7–14 Hz (Jones, 2002; Crabtree, 2018). Our model postulates a different mechanism for generating alpha rhythms from the balanced *E-I* networks' intrinsic properties (see above) and the oscillatory property evoked by the interplay of short-term plasticity components. The combination of facilitatory and depression synapses can generate time-dependent connectivity effects, which engage pulvinar neurons to oscillate in specific low-band frequencies. Such resonant activation acts as a band-pass filter, eliciting a specific amount of neuronal information and communication between the cortex and the thalamus (Markram et al., 1998; O'Brien et al., 2014). The model, parameterized with the experimental proportion of CT terminal types 1 and 2 of two cortical areas, suggests a possible non-GABA mediated mechanism as suggested by recent pulvinar causal manipulation results in monkey (Saalmann et al., 2012; Zhou et al., 2016).

In a more natural scenario, the addition of noise and other membrane dynamics favored the appearance of more realistic scenarios similar to previous experimental works. The addition of a synaptic noise background (**Figure 9A**) allowed decreasing the delay to reach this oscillatory stability (Silver, 2010). The addition of burst-like dynamics favored irregular oscillations and lower power in the amplitude of the signal (Ramcharan et al., 2005; O'Reilly et al., 2021). However, when the cortico-pulvino-cortical model was implemented, we selected pulvinar neurons with regular spiking responses with a low background synaptic noise (case A, **Figures 10, 11**). Such properties were chosen to reinforce the regular oscillatory activity of the pulvinar on the neural responses of the visual cortex. Although the burst-like responses may refine the neuronal responses of the pulvinar to a more realistic scenario, our model predicts that such stable oscillatory responses are due to the physiological properties of terminal types 1 and 2 and how these are combined to obtain a nearly balanced state.

Implications in the Transmission of Oscillatory Cortical Responses Throughout the Transthalamic Pathway

Visual processing in cortical areas follows mostly a hierarchical stream of communication, in which the visual information travels from low to high levels (feedforward pathway), as well as from high to low levels (feedback pathway) (Felleman and Van Essen, 1991; Bullier, 2003). Oscillatory-band activity has been associated with these anatomical pathways, in which gamma-band oscillations are found in the feedforward route, whereas slow oscillations, alpha/beta waves, are observed in the feedback direction (van Kerkoerle et al., 2014; Bastos et al., 2015). In our simulations, nonetheless, we show that the alpha-type oscillatory activity evoked in area 21a, we are able to reverse that transmission in feedback action if $\eta^{21a}_1 > \eta^{21a}_2$, *so* reversing the ratio of types 1 and 2 terminals, and introducing a pulvino-cortical projection such as $W_{CP}{}^{a17} < 0$. In this new scenario, CT projections from area 21a would drive alpha-band oscillatory responses in the pulvinar, whereas CT connections from area 17 to the pulvinar would shift such responses into asynchronous-type activity (**Figure 13**).

Reversing the direction of transmission along the cortical hierarchy, and changing the activity from a synchronized to an unsynchronized state, depends on two properties of the simulated network. As described above, terminal types 1 and 2 can evoke alpha-band oscillations independently. Interestingly, as the terminals are set with the ratios of areas 17 and 21a, solutions are found when the predominant terminal had a higher weight of connection than the other type. That is, to engage alpha waves in the pulvinar by CT projections from area 17, the type 2 terminal had to be stronger than the type 1, and vice versa for area 21a, which follows the ratios found in the empirical anatomical data. The second property is derived from the two compartments organized in the pulvinar as striate- and extrastriate-recipient zones. Our model predicts that activity between these zones would compete with each other for activation in the pulvinar, similar to a "Winner-Take all" (WTA) process. Since these zones are activated independently by different cortical areas along the visual hierarchy, asynchronous or synchronous activity within the pulvinar will also be triggered independently. Such hypothetical compartmentalization of responses predicted by the model allows the pulvinar to revert its oscillatory alpha-band activity to an asynchronous irregular spiking or vice versa as shown by recent experimental data (see below). Thus, the pulvinar filters one type of neural response over the other, depending on which cortical area projects more strongly to it.

The pulvinar, across different species, is a heterogeneous structure that is composed of multiple subdivisions and, unlike the LGN, such separations have little organizational arrangements like neuronal lamellae (Baldwin and Bourne, 2017). These anatomical features mean that defining the function of the pulvinar throughout mammalian evolution remains a challenge (Casanova, 2003). Our model, with at least two-separated cortical-recipient zones, may clarify functional aspects that the pulvinar has. As our model predicts, different low-band frequencies are built in the pulvinar as the strength of CT connections increases gradually, mainly when type 1 connections are used (**Figures 4**, **11**). If several cortical areas are represented independently as separate connections domains in the pulvinar (Shipp, 2003), these pulvinar domains could generate different oscillations regulating visual cortical activity and coordinating transthalamic messages in an oscillatory ascending manner to the visual cortex. Thus, like a WTA computation, the pulvinar would serve as a channel selector of different band frequencies that would adjust cortical dynamics to be able to transmit oscillatory low-frequency activity from one group of neurons to another, possibly in a feedforward or feedback manner throughout the visual hierarchy (Quax et al., 2017; Jaramillo et al., 2019; Cortes et al., 2020). In other words, the pulvinar could select and separate signals from and to the visual cortex by the above-described WTA mechanism. According to this view, the pulvinar would use two anatomical properties to allow visual processing in the cortex: a gradient of terminal types 1 and 2 throughout the cortical hierarchy (Abbas-Farishta et al., 2020), and the spatial processing by the compartmentalization of cortical-recipient zones suggested by our model. Therefore, the pulvinar would need a differential increase in type 1 terminal weights along higher order visual cortical areas and cortico-recipient zones that are partially isolated to carry out this selection mechanism (Huppe-Gourgues et al., 2006, 2019; Abbas-Farishta et al., 2020).

The addition of more inputs into the pulvinar would explain the appearance of other low-frequency oscillations observed in *in vivo* experiments. As our simulations predict, the increase of strengths of the CT projection from area 21a (**Figures 3**, **11**) causes a shift to other oscillatory frequencies by the competition between rival cortical-recipient zones in the pulvinar. One interpretation of these results may be the recruitment of other higher cortical areas increasing the number of CT type 1 endings along the visual cortex that reach and activate synergistically the pulvinar. According to this view, η_1 in our simulations is not static; it changes dynamically throughout the visual cortical hierarchy to meet behavioral demands matching cognitive states (i.e., attentional demands). As more and more areas project to the pulvinar, this top-down activation could generate different band frequencies, which is seen for theta-, beta- and gamma-band oscillations in awake animals (Wrobel et al., 2007; Saalmann and Kastner, 2011; Yu et al., 2018). On the one hand, the pulvinar may maintain specific resonant frequencies with specific cortical areas that allow cortico-cortical feedforward and feedback processing to propagate in one direction or the other (Saalmann et al., 2012; Zhou et al., 2016; Cortes et al., 2020). Another simple interpretation of alpha rhythms in the pulvinar are not optimized for coupling between cortical areas, as areas 17 and 21a, since one cortical area would dominate its representation inside the pulvinar over the others. Future experiments will reveal what mechanism the pulvinar uses to maintain its functional coherence with the visual cortex.

On the other hand, the oscillatory response provoked by the joint action of the two areas (**Figure 8**) could also be interpreted as a convergent action of connections coming from the visual cortex and other brain structures. As our model shows, we found a solution to evoke alpha rhythms when areas 17 and 21a had similar weights of connections. Since the connectivity weights of

the cortical areas across the visual hierarchy are unlikely to be similar (because of their different physiological characteristics), one prediction of the model is that one of those connections is not from the visual cortex. Instead, it would arise from subcortical projections, for example, from the superior colliculus. This joint cortical and subcortical action could engage pulvinar responses in different oscillatory modes (Le et al., 2019), when, for example, eye movements are required (Berman and Wurtz, 2011). Thus, the recruitment of other cortical areas and the combination with other subcortical structures could explain the different oscillatory ranges found in the pulvinar.

Another prediction of our model and the WTA proposed mechanism is that the inhibition of one compartment in the pulvinar can increase the activity in another cortico-recipient zone. For example, in **Figure 12**, stopping the extrastriate-recipient zone's activity produces excitation of the striate-recipient zone, which causes an increase of alpha-band oscillations in area 21a by the transthalamic pathway. This excitatory oscillatory effect on cortical populations of neurons when local inhibitions are made in restricted areas of the pulvinar has been already quantified experimentally (Cortes et al., 2020). Alternatively, this antagonistic effect could also explain the effect of GABA inactivation in the pulvinar, in which the firing rate of cortical neurons achieves a reduction and an enhancement in areas 17 and 21a of cats, respectively (de Souza et al., 2020). Likewise, pulvinar inactivation potentiates both stimulus-driven responses in monkey V2 cells (Soares et al., 2004) and low-frequency LFP power during monkey visual attentional tasks (Zhou et al., 2016). Results that point in the same direction that the suppressive effects of the surrounding regions of V1 receptive fields when pulvinar is excited (Purushothaman et al., 2012). The suggested compartmentalization of the pulvinar would explain why disruption of the activity of isolated domains could create excitation of other nearby domains within the pulvinar.

Implications of Pulvinar Bistable States

Perception of low contrast stimuli reveals that the brain shows two temporal states of visual awareness processing (van Vugt et al., 2018). The brain processes external stimuli in a modular and parallel bottom-up hierarchical fashion if the stimulus is strong enough and exceeds an internal threshold of visual perception. This first state is essentially feedforward. Such signals are retrieved and selected in a second state through attentional requirements located in hierarchically elevated cortical areas. Here, the flow pathway is in a feedback manner, favoring convergence to activate necessarily and sufficiently all the network nodes (Dehaene et al., 2011).

Although both cortical states have been explained based on the long-range excitatory cortical connection, we postulate that the pulvinar is involved in changing from one state to another. Such control would be possible because of the bi-state generated by the hierarchical gradient of CT terminal types 1 and 2 reaching the pulvinar. Theoretically, it has been shown that the pulvinar and its transthalamic pathway are necessary to pass neural responses in a graded manner through a chain of sequentially connected areas (Cortes and van Vreeswijk, 2012, 2015; de Souza et al., 2020). Such a prediction

of the functionality of the transthalamic pathway has been partially corroborated by experiments in the visual system of cats (de Souza et al., 2020). On the other hand, the differential oscillatory response of the neuronal signals between the cortices in certain phases of pulvinar oscillatory activity (Saalmann and Kastner, 2011; Saalmann et al., 2012; Fiebelkorn et al., 2019), as well as when the pulvinar is inactivated (Zhou et al., 2016; Cortes et al., 2020), suggest it plays a role in the effective cortical connectivity. Our results showing bi-stable pulvinar states, suggest that cortical feedback transmission is associated with pulvinar oscillatory activity and the feedforward pathway, with the asynchronous irregular spike response of the pulvinar. Thus, this parallel feedback pathway through the pulvinar would reinforce the top-down attentional activity associated with low-frequency oscillations (*i.e.,* areas 17 and 18). This prediction of feedforward oscillatory activation should be verified in future works, since pulvinar could be indirectly activated by cortico-cortical feedback. That is, from layers 2/3 of a higher cortical area to layer 5 of a lower cortical area, and from here to the pulvinar (e.g., from area 21a to area 17). Such a feedback pathway would induce the type of antagonistic dynamics predicted here (**Figure 13**).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

NC, RA, and HL: conceptualization. NC: methodology, software, formal analysis, investigation, data curation, and writing—original draft preparation. NC and CC: validation and visualization. CC: resources, supervision, project administration, and funding acquisition. NC, RA, HL, and CC: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncel.2021.787170/full#supplementary-material

6. Beyond V1: Variance and Thalamo-Cortical Loops – 6.4. Article: "Corticothalamic Projections Gate Alpha Rhythms in the Pulvinar"

REFERENCES

Abbas-Farishta, R., Boire, R., and Casanova, C. (2020). Hierarchical organization of corticothalamicprojections to the pulvinar. *Cereb. Cortex* 1:tgaa030.

Abbas-Farishta, R., Zouahi, H., and Casanova, C. (2021). Distributions of vesicular glutamate transporters 1 and 2 in the visual thalamus and associated areas of the cat. *J. Comput. Neurol.* [Epub ahead of print], doi: 10.1002/cne.25239

Baldwin, M. K. L., and Bourne, J. A. (2017). "The evolution of subcortical pathways to the extrastriate cortex," in *Evolution of Nervous Systems*, 2nd Edn, ed. J. H. Kaas (Elsevier), 165–185. doi: 10.1016/b978-0-12-804042-3.00081-6

Bastos, A. M., Donoghue, J. A., Brincat, S. L., Mahnke, M., Yanar, J., Correa, J., et al. (2021). Neural effects of propofol-induced unconsciousness and its reversal using thalamic stimulation. *Elife* 10:e60824. doi: 10.7554/eLife.60824

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J. M., Oostenveld, R., Dowdall, J. R., et al. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* 85, 390–401. doi: 10.1016/j.neuron.2014.12.018

Bender, D. B. (1983). Visual activation of neurons in the primate pulvinar depends on cortex but not colliculus. *Brain Res.* 279, 258–261. doi: 10.1016/0006-8993(83)90188-9

Berman, R. A., and Wurtz, R. H. (2011). Signals conveyed in the pulvinar pathway from superior colliculus to cortical area mt. *J. Neurosci.* 31, 373–384. doi: 10.1523/jneurosci.4738-10.2011

Brette, R., and Gerstner, W. (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *J. Neurophysiol.* 94, 3637–3642. doi: 10.1152/jn.00686.2005

Bullier, J. (2001). Integrated model of visual processing. *Brain Res. Rev.* 36, 96–107. doi: 10.1016/s0165-0173(01)00085-6

Bullier, J. (2003). "Hierarchies of cortical areas," in *The Primate Visual System*, eds J. Kaas and C. Collins (Boca Raton, FL: CRC Press), 181–204.

Casanova, C. (2003). "The visual functions of the pulvinar," in *The Visual Neurosciences*, eds L. Chalupa and J. Werner (The MIT Press), 592–680.

Casanova, C. (2021). "Pulvinar and transthalamic cortical visual pathway," in *The Senses: A Comprehensive Reference*, 2nd Edn, eds R. H. Masland, T. D. Albright, and E. P. Gardner (Elsevier), 472–479. doi: 10.1016/b978-0-12-809324-5.23899-6

Chalupa, L. M. (1991). "The neural basis of visual function," in *Vision and Visual Dysfunctions*, eds B. Dreher and S. R. Robinson (CRC Press, Inc.), 140–159.

Cortes, N., de Souza, B., and Casanova, C. (2020). Pulvinar modulates synchrony across visual cortical areas. *Vision* 4, 1–18. doi: 10.3390/vision4020022

Cortes, N., and van Vreeswijk, C. (2012). The role of pulvinar in the transmission of information in the visual hierarchy. *Front. Comput. Neurosci.* 6:1–21. doi: 10.3389/fncom.2012.00029

Cortes, N., and van Vreeswijk, C. (2015). Pulvinar thalamic nucleus allows for asynchronous spike propagation through the cortex. *Front. Comput. Neurosci.* 9:60. doi: 10.3389/fncom.2015.00060

Crabtree, J. W. (2018). Functional diversity of thalamic reticular subnetworks. *Front. Syst. Neurosci.* 12:41. doi: 10.3389/fnsys.2018.00041

de Souza, B., Cortes, N., and Casanova, C. (2020). Pulvinar modulates contrast responses in the visual cortex as a function of cortical hierarchy. *Cereb. Cortex* 30, 1068–1086. doi: 10.1093/cercor/bhz149

de Souza, B., Frigon, E., Tremblay-Laliberté, R., Casanova, C., and Boire, D. (2021). Laminar distribution of cortical projection neurons to the pulvinar: a comparative study in cats and mice. *J. Comput. Neurol.* 529, 2055–2069. doi: 10.1002/cne.25072

Dehaene, S., Changeux, J., and Naccache, L. (2011). *The Global Neuronal Workspace Model of Conscious Access: From Neuronal Architectures to Clinical Applications*. Berlin: Springer Berlin Heidelberg, 55–84.

Feig, S., and Harting, J. (1998). Corticocortical communication via the thalamus: ultrastructural studies of corticothalamic projections from area 17 to the lateral posterior nucleus of the cat and inferior pulvinar nucleus of the owl monkey. *J. Comput. Neurol.* 395, 281–298. doi: 10.1002/(sici)1096-9861(19980808)395:3<281::aid-cne2>3.0.co;2-z

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebr. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1

Fiebelkorn, I., Pinsk, M., and Kastner, S. (2019). The mediodorsal pulvinar coordinates the macaque fronto-parietal network during rhythmic spatial attention. *Nat. Commun.* 10:215. doi: 10.1038/s41467-018-08151-4

Halgren, M., Ulbert, I., Bastuji, H., Fabó, D., Eröss, L., Rey, M., et al. (2019). The generation and propagation of the human alpha rhythm. *Proc. Natl. Acad. Sci. U.S.A.* 116, 23772–23782.

Hegde, J., and Felleman, D. J. (2007). Reappraising the functional implications of the primate visual anatomical hierarchy. *Neuroscientist* 13, 416–421.

Huppe-Gourgues, F., Abbas-Farishta, R., Boire, D., Ptito, M., and Casanova, C. (2019). Distribution and morphology of cortical terminals in the cat thalamus from the anterior ectosylvian sulcus. *Sci. Rep.* 9:3075.

Huppe-Gourgues, F., Bickford, M., Boire, D., Ptito, M., and Casanova, C. (2006). Distribution, morphology, and synaptic targets of corticothalamic terminals in the cat lateral posterior-pulvinar complex that originate from the posteromedial lateral suprasylvian cortex. *J. Comput. Neurol.* 497, 847–863. doi: 10.1002/cne.21024

Izhikevich, E. M., Desai, N. S., Walcott, E. C., and Hoppensteadt, F. C. (2003). Bursts as a unit of neural information: selective communication via resonance. *Trends Neurosci.* 26, 161–167. doi: 10.1016/S0166-2236(03)00034-1

Jaramillo, J., Mejias, J. F., and Wang, X. J. (2019). Engagement of pulvino-cortical feedforward and feedback pathways in cognitive computations. *Neuron* 101, 321–336. doi: 10.1016/j.neuron.2018.11.023

Jones, E. G. (2002). Thalamic circuitry and thalamocortical synchrony. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 357, 1659–1673. doi: 10.1098/rstb.2002.1168

Le, Q. V., Nishimaru, H., Matsumoto, J., Takamura, Y., Nguyen, M. N., Mao, C. V., et al. (2019). Gamma oscillations in the superior colliculus and pulvinar in response to faces support discrimination performance in monkeys. *Neuropsychologia* 128, 87–95. doi: 10.1016/j.neuropsychologia.2017.10.015

Li, J., Guido, W., and Bickford, M. (2003). Two distinct types of corticothalamic epsps and their contribution to short-term synaptic plasticity. *J. Neurophysiol.* 90, 3429–3440. doi: 10.1152/jn.00456.2003

Li, K., Patel, J., Purushothaman, G., Marion, R., and Casagrande, V. (2013). Retinotopic maps in the pulvinar of bush baby (otolemur garnettii). *J. Comput. Neurol.* 521, 3432–3450. doi: 10.1002/cne.23358

Markram, H., Wang, Y., and Tsodyks, M. (1998). Differential signaling via the same axon of neocortical pyramidal neurons. *Proc. Natl. Acad. Sci. U.S.A.* 95, 5323–5328. doi: 10.1073/pnas.95.9.5323

Nassi, J. J., and Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nat. Rev. Neurosci.* 10, 360–372. doi: 10.1038/nrn2619

O'Brien, M. J., Thibeault, C. M., and Srinivasa, N. (2014). A novel analytical characterization for short-term plasticity parameters in spiking neural networks. *Front. Comput. Neurosci.* 8:148.

Ojima, H., Murakami, K., and Kishi, K. (1996). Dual termination modes of corticothalamic fibers originating from pyramids of layers 5 and 6 in cat visual cortical area 17. *Neurosci. Lett.* 208, 57–60. doi: 10.1016/0304-3940(96)12538-6

O'Reilly, R., Russin, J., Zolfaghar, M., and Rohrlich, J. (2021). Deep predictive learning in neocortex and pulvinar. *J. Cogn. Neurosci.* 33, 1158–1196. doi: 10.1162/jocn_a_01708

Payne, B. (1993). Gating and control of primary visual cortex by pulvinar. *Cereb. Cortex* 3, 1–25.

Purushothaman, G., Marion, R., Li, K., and Casagrande, V. (2012). Gating and control of primary visual cortex by pulvinar. *Nat. Neurosci.* 15, 905–912. doi: 10.1038/nn.3106

Quax, S., Jensen, O., and Tiesinga, P. (2017). Top-down control of cortical gamma-band communication via pulvinar induced phase shifts in the alpha rhythm. *PLoS Comput. Biol.* 12:e1005519. doi: 10.1371/journal.pcbi.1005519

Ramcharan, E., Gnadt, J., and Sherman, S. (2005). Higher-order thalamic relays burst more than first-order relays. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12236–12241. doi: 10.1073/pnas.0502843102

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., et al. (2010). The asynchronous state in cortical circuits. *Science* 327, 587–590. doi: 10.1126/science.1179850

Rockland, K. S. (2019). Corticothalamic axon morphologies and network architecture. *Eur. J. Neurosci.* 49, 969–977. doi: 10.1111/ejn.13910

Rodney, D., Markram, H., and Martin, K. (2004). "Neocortex," in *The Synaptic Organization of the Brain*, ed. G. M. Shepherd (Oxford University Press), 449–558.

158

Saalmann, Y. B., and Kastner, S. (2011). Cognitive and perceptual functions of the visual thalamus. *Neuron* 71, 209–223. doi: 10.1016/j.neuron.2011.06.027

Saalmann, Y. B., Pinsk, M. A., Wang, L., Li, X., and Kastner, S. (2012). The pulvinar regulates information transmission between cortical areas based on attention demands. *Science* 337, 753–756. doi: 10.1126/science.1223082

Sherman, S., and Guillery, R. (2002). The role of the thalamus in the flow of information to the cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 357, 1695–1708. doi: 10.1098/rstb.2002.1161

Sherman, S. M., and Guillery, R. W. (2011). Distinct functions for direct and transthalamic corticocortical connections. *J. Neurophysiol.* 106, 1068–1077. doi: 10.1152/jn.00429.2011

Shipp, S. (2003). The functional logic of cortico-pulvinar connections. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 358, 1605–1624. doi: 10.1098/rstb.2002.1213

Shipp, S. (2004). The brain circuitry of attention. *Trends Cogn. Sci.* 8, 223–230. doi: 10.1016/j.tics.2004.03.004

Silver, R. (2010). Neuronal arithmetic. *Nat. Rev. Neurosci.* 11, 474–489. doi: 10.1038/nrn2864

Snow, J. C., Allen, H. A., Rafal, R. D., and Humphreys, G. W. (2009). Impaired attentional selection following lesions to human pulvinar: evidence for homology between human and monkey. *Proc. Natl. Acad. Sci. U.S.A.* 106, 4054–4059. doi: 10.1073/pnas.0810086106

Soares, J., Diogo, A., Fiorani, M., Souza, A., and Gattass, R. (2004). Effects of inactivation of the lateral pulvinar on response properties of second visual area cells in cebus monkeys. *Clin. Exp. Pharmacol. Physiol.* 31, 580–590. doi: 10.1111/j.1440-1681.2004.04051.x

Stimberg, M., Goodman, D., Brette, R., and Pittà, M. (2019b). "Modeling neuron-glia interactions with the brian 2 simulator," in *Springer Series in Computational Neuroscience*, eds M. De Pittà and H. Berry (Cham: Springer), 471–505. doi: 10.1007/978-3-030-00817-8_18

Stimberg, M., Brette, R., and Goodman, D. F. (2019a). Brian 2, an intuitive and efficient neural simulator. *eLife* 8:e47314.

Tsodyks, M., Pawelzik, K., and Markram, H. (1998). Neural networks with dynamic synapses. *Neural. Comput.* 10, 821–835. doi: 10.1162/08997669830001 7502

van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., van der Togt, C., et al. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14332–14341. doi: 10.1073/pnas.1402773111

van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274, 1724–1726. doi: 10.1126/science.274.5293.1724

van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., et al. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science* 360, 537–542. doi: 10.1126/science.aar 7186

Vidnyánszky, Z., Borostyánköi, Z., Görcs, T., and Hámori, J. (1994). Light and electron microscopic analysis of synaptic input from cortical area 17 to the lateral posterior nucleus in cats. *Exp. Brain Res.* 109, 63–70. doi: 10.1007/BF00228627

Vogels, T., and Abbott, L. (2009). Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nat. Neurosci.* 12, 483–491. doi: 10.1038/nn.2276

Wilke, M., Mueller, K. M., and Leopold, D. A. (2009). Neural activity in the visual thalamus reflects perceptual suppression. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9465–9470. doi: 10.1073/pnas.0900714106

Wimborne, B., and Henry, G. (1992). Response characteristics of the cells of cortical area 21a of the cat with special reference to orientation specificity. *J. Physiol.* 449, 457–478. doi: 10.1113/jphysiol.1992.sp019096

Wrobel, A., Ghazaryan, A., Bekisz, M., Bogdan, W., and Kaminski, J. (2007). Two streams of attention-dependent beta activity in the striate recipient zone of cat's lateral posterior-pulvinar complex. *J. Neurosci.* 27, 2230–2240. doi: 10.1523/JNEUROSCI.4004-06.2007

Yu, C., Li, Y., Stitt, I., Zhou, Z., Sellers, K., and Frohlich, F. (2018). Theta oscillations organize spiking activity in higher-order visual thalamus during sustained attention. *ENeuro* 5:ENEURO.0384-17.2018. doi: 10.1523/ENEURO.0384-17.2018

Zhou, H., Schafer, R. J., and Desimone, R. (2016). Pulvinar-cortex interactions in vision and attention. *Neuron* 89, 209–220.

159

6.5. Conclusion

In this section, we explored the complex functional relationship between the pulvinar and V1, with an emphasis on the regulatory processes associated with inverse variance computations.

The anatomy of the pulvinar seems to be uniquely well-posed for inverse variance modulations in the visual hierarchy. Indeed, the functional connectivity is strictly modulatory from pulvinar to V1, meaning that transthalamic connections could regulate prediction errors based on inverse variance right at the lowest level of the hierarchy. This modulatory activity likely has the role of preventing imprecise sensory inputs from propagating in the hierarchy, thus preventing them from updating (wrongly) an internal predictive model due to unreliable sensory input, which is very fitting with the gating role commonly assigned to the pulvinar [252]. Further, the asymmetrical nature of this connectivity between facilitating (type 1) and depressing (type 2) synapses, established as a function of the cortical hierarchy, fits very nicely with hierarchical Bayesian inference. In practice, this means that the pulvinar can compute a global associative signal linked to multiple level of inverse variance through the visual areas, and explain away irrelevant visual features. Deficit in this role, intuitively, can be conceptualized as propagation of irrelevant prior predictions, which is often reported in pulvinar lesion studies [258].

The propagation of precision-weighted prediction and errors is also likely supported by neuromodulatory mechanisms that control synaptic gain, namely cholinergic signals [215]. As with the pulvinar, this mechanism provides a way to encode information related to environment through the excitability of circuits signalling prediction errors [162]. In line with our results, this further aligns very well with the idea that superficial pyramidal cells are equipped with numerous synaptic gain control systems neuromodulatory receptors. Further, the clear effect of neuromodulatory mechanisms on synchronous activity in the cortex, and on alpha oscillations, hints at an interconnected mechanism throughout multiple neuronal levels [39].

Including such transthalamic and neuromodulatory connectivity in our view of the cortex is an important consideration, and this section offered a brief divergence to add some much needed mesoscale context to our results. We will now revisit, in the concluding section of this manuscript, cortical-centric computations enriched by an understanding of the hierarchical interplay afforded by the pulvinar. We posit that the pulvinar's interaction with cortical regions represents a second-order mechanism for implementing inverse variance computations in the brain, a theme we will expand upon when synthesizing the multiscale findings in the next chapter.

7. Conclusion

"Well, I'm all for leaving
and that being done,
I've put in a request to take up my turn."
Jethro Tull, A Passion Play, 1973

7.1. Concluding Overview

7.1.1. Final Summary

As the author of this manuscript easily recognized during proofreading, sifting through more than two hundred pages of early-career neuroscience is not exactly a straightforward ordeal. Although an attempt has been made to lower the attentional overhead involved in this reading by humoring the reader and by limiting the number of abbreviations, it is likely that days have elapsed between the beginning and the end of one's reading. As such, we would like to summarize, once last time, this manuscript

We first introduced the problem of inverse variance computations in vision by adopting a Marr-like approach in chapter 2. This introduced a study of the nature of natural images in chapter 3, exploring the trade-off between fidelity and sparsity, and their respective advantage when computing variance-bound probability distributions. This concept was pursued in chapter 4 in an electrophysiological study, leading to major contributions in terms of the functional role of cortical recurrence in computations of variance. We then built further on these concepts in chapter 5, introducing cyclic computational graphs and reinforcing the idea of a multiscale model of orientation selectivity. This was then extended in chapter 6, focusing beyond V1 and delves into the subcortical pathways, particularly highlighting the role of the pulvinar in managing variance in visual processing. Preliminary results from extrastriate cortical areas also hint at their potential involvement in variance computations. Now, in chapter 7, we propose a (short) reflective summary, pondering the broad implications of these results for neuroscience and related fields. If the reader has any courage left, they are also provided with optional appendices that provide a deeper dive into the mathematical underpinnings of the free energy principle, as introduced at the beginning of this manuscript. The last appendix also includes supplementary material related to public communications surrounding this thesis.

7.1.2. Towards a Coherent Model of inverse variance-weighting

By nature, this manuscript encompasses a broad number of topics, including sparse coding, extracellular electrophysiology, artificial deep neural networks, graph theory, and thalamocortical communication. Rather than subject of studies in themselves, these were more designed as means towards an end - or rather, as hammers for one single nail (maybe spike, in our context). By giving coherent and convergent observation of the computations of inverse variance as based on recurrence, these contributions must now be framed in a single coherent framework.

Based on our results, we propose the following. The computation of inverse variance starts with the notion that the goal of primary sensory areas is to extract invariant representations of the world [163, 18]. This might seem semantically contradictory, but if representations are to be invariant, then they must be encoded alongside an associated variance in the brain's internal models. Both under predictive coding or sparse coding, the goal of these internal models system is to build a representation of the world whilst enforcing the cost/efficiency tradeoff. In chapter 3, we show that the low-level invariant representations of natural images are oriented edges with factorized variance. A model of the visual world can rely on two strategies: encoding a distribution of edges using multiple single units (accurate, but not sparse) or encoding a distribution of edges using one unit, that encodes both median and variance (not necessarily accurate, but sparse). This latter scheme seems to be used, and can be likened to a Maximum Likelihood Estimation in the context of the Bayesian Brain (see Equation 2.6). In chapter 4, we show that both strategies are likely implemented in the neural substrate: many neurons that encode single edges, through a fast first pass in vulnerable neurons, a representation which then becomes a sparse but approximate estimate of variance in resilient neurons. In chapter 5, we also view this as a dynamically modulated (by variance) message passing through many neurons carrying different orientation information, creating a local, recurrence-based, strategy for encoding orientation variance.

Given that this computation can be created solely by local intra-area recurrence, it's likely running locally in parallel in every single cortical area. Quite probably, it is also running in every cortical area, visual or not. Thus, there must be a system of communication between these local processes that allows to regulate the flow of information based on these multiple inverse variance computations. Having a central modulator of these computations, in the form of the pulvinar, which controls propagation of (alpha-oscillating) predictions as shown in chapter 6, proves to be computationally advantageous and conceptually elegant. The longer time constants of this large-scale thalamocortical networks can also implement the computation of inverse variance through time, as discussed in chapter 6, which is also supported by empirical evidences of attention-deficit due to pulvinar lesions. It is likely that neuromodulator mechanisms are also involved [215], which could be the brain-wide equivalent of the role the pulvinar is fulfilling for the visual cortex. This constitutes our main scientific contribution, which, to quote the title of this thesis, is *"A multiscale*

model to account for orientation selectivity in natural images".

This conceptual model also speaks to a number of testable hypotheses that can pave the way for exciting future research perspectives. But first, let us review the possible interpretations of this work, under both at the neurobiological and the computational front.

7.2. Interpretation

Although neuroscience has been historically based on studies related to how neurons encode only scalar approximate representations of their environment [281, 137, 229], theoretical [18] and experimental [189] convergences have made it possible to start considering the full statistical nature of the environment we live in. The current research on inverse variance-weighting in the brain has exciting fundamental and clinical perspectives, which relates to two major avenues of interpretations.

7.2.1. Neurobiological Relevance

At its core, inverse variance-weighting addresses the essential balance between internal priors and external perceptions. In the context of predictive processing, this mechanism guides the brain in discerning which of its predictions are reliable, and which it should discard in favor of prediction errors which warrant new attention. Interestingly, this dynamic can be seen as a foundational descriptor for several psychiatric disorders [95], including autism and schizophrenia.

First, on a clinical level, recall that in the introduction (Figure 2.10), we discussed the notion that the variance of either prediction or prediction errors could drive the posterior likelihood towards one or the other. Under that generic description, hypersensitivity to external inputs, as is present in autistic spectrum disorders, can be conceptualized as a condition that drives overly precise prediction errors, overriding any possible internal models of the world. Typically, overlooking prediction error with no consequence for our inner model is something that we all do on a daily basis. In complex but non-threatening scenarios like social interactions, some errors are acceptable for broader understanding, and are part of the human social basis for learning. Inflexible processing of prediction errors in autism, due to impossibility of lowering their inverse variance-weighting in the decision process, leads to inability in ensuring "predictive success in an unpredictable world" [324].

On the other end of the possible symptoms captured by inverse variance models lies schizophrenia. Hallucinatory experiences and feeling of disconnection with the external world [142] are the hallmark of high weighting given to internal models over external likelihoods [299]. This is further supported by evidences that many of these symptoms align closely with the lesions of the neural substrate we described here, notably, visual perception and visual cognition being affected by pulvinar lesions [71]. Functional connectivity in schizophrenic patients have revealed reduced connectivity between the medial pulvinar and the frontal cortex [242, 343]. In the general model
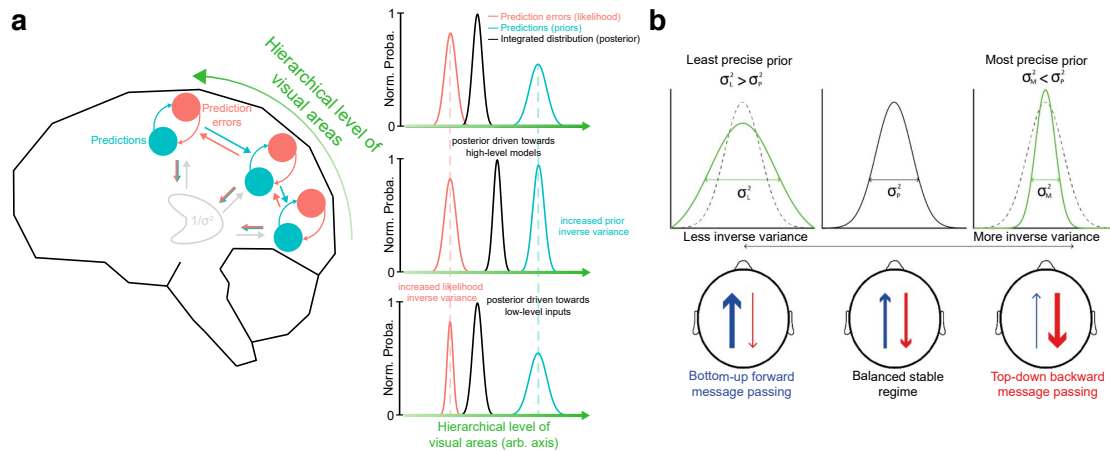
Figure 7.1. – Computational psychiatry under predictive processing. (a) Message passing in the hierarchical Bayesian visual cortex. Under abnormally high inverse variance of the prior (second row), posterior is driven towards the internal models. Oppositely, high inverse variance of prediction errors (third row), drives posterior towards the sensory input. These respectively account for positive symptoms in schizophrenia and autism. (b) Same as (a), but in terms of bottom-up and top-down message passing, reproduced from [8].

proposed earlier in this thesis, this would mean a compromised ability to regulate the flow of neuronal information between cortical areas by means of inverse variance weighting. Conversely, it could provide an evolutionary basis for the need for a central structure involved in the regulation of this mechanism, with a size proportional to that of the cortical areas being regulated [46].

Second, on a cognitive level, another contribution of this thesis is to also offer a shift of our understanding of attention mechanisms, envisioning them as inverse variance mechanisms. Instead of conceptualizing recurrence and lateral inhibition as neural basis for sensory sharpening, one could consider that these are instead sampling mechanisms to drive (local) emphasis on the optimal percepts. Under this view, an interesting link is also to be made with the neuromodulatory activity, classically related to attention. Dopamine, for instance, is suggested to be crucial for the precision-adjustment of prediction errors, which in turn signals the significance of sensory inputs [89, 274]. Other neuromodulatory mechanism, such as cholinergic systems, can increase the gain of cortical neurons in response to sensory inputs [327, 38], and are thus tied to attentional mechanisms [155]. In that sense, these systems could be thought of as a third-order implementation of modulation by inverse variance, right after the first-order local recurrent interactions, second-order vision-centric pulvinar pathways, and serving to implement diffuse cortex-wise computations of inverse variance.

Third, on a fundamental level, the dynamics of the activities that have been examined so far, as well as their propagation methods, their intracortical nature, all point

towards an exciting research perspective for travelling waves. As introduced in chapter 4, traveling waves represent the dominant way of propagation of activity within the cortex. These waves facilitate processes like priming and suppression, but their exact role remains somewhat ambiguous and polymodal [217]. Similarly to the idea of a cholinergic mechanism, travelling waves also modulate the excitation/inhibition balance of the cortical activity rhythmically, which gates perceptual inference [75]. At the cortical area level, one contribution of this thesis is to consider that these waves might also implement a local inverse variance-weighting phenomenon, which offers a fresh perspective on their function in the cortex. At the macroscale, waves travelling between cortical areas, could also reveal mechanisms of propagation of inverse variance-weighted message, controlled by a series of locally weighted predictive oscillators [9]. Specific changes in inverse variance could lower the excitability of prediction error units in the lower end of the hierarchy (i.e., in V1), which would increase the preponderance of backward traveling waves, and correlates nicely with attentional experiments [9]. One limitation here is that we have chosen to limit ourselves to an alpha/gamma duo, which is an oversimplification, and does not consider the fact that attentional visual search tasks correlates with even lower frequency rhythms, like theta [275, 209]. Given the tight relationship between unpredictable environment and the need for attentive search in feature space, this offers an interesting experimental perspective on manipulating visual variance, and correlating neural activity at the scale of the brain with the need to drive away - or close to - the input likelihood from posterior distributions. This will be discussed further in the final section of this manuscript.

7.2.2. Machine Learning Relevance

Inverse variance-weighting plays a significant role in both statistical and machine learning research [219]. Under this interpretation, inverse variance provides distinct advantages to a model, without implementational drawbacks (as studied in the neuroscientific context here). Indeed, given the mathematical triviality to compute a correlational matrix and inverse it, modelers obtain all the theoretical advantages described so far, with none of the complications related to recurrent neural message passing. One key advantage of implementing a degree of variance to a model's decision is to allow the readout of the model's confidence in its output. This can be as straightforward, in a deep neural network, as computing the sum of the variances in all layers, such that a network that has learned a precise representation of the environment returns a globally low variance of its activity.

Recently, the importance of associating confidence levels with algorithmic decisions has grown, placing inverse variance-weighting at the forefront of machine learning algorithm development. While there might be lighthearted examples, such as deep learning models mistakenly classifying an image of a panda with a single modified pixel as a dog, the stakes are much higher in practical industrial applications. As the research shifts towards implementing safety-critical neural networks, the emphasis on variance-weighted decisions becomes paramount. To quote Kendall and Yarin [164]:

"Mappings are often taken blindly and assumed to be accurate, which is not always the case. In (. . .) recent examples, this has had disastrous consequences. In May 2016 there was the first fatality from an assisted driving system, caused by the perception system confusing the white side of a trailer for bright sky. If (. . .) these algorithms were able to assign a high level of uncertainty to their erroneous predictions, then the system may have been able to make better decisions and likely avoid disaster."
from Alex Kendall and Yarin Gal. "What uncertainties do we need in bayesian deep learning for computer vision?" In: *Advances in neural information processing systems* 30 (2017)

Incorporating explicit learning rules based on variance into data modeling undeniably introduces additional complexity. As we have shown in the introduction of this manuscript and in the conclusion of chapter 5, these merely require simple additional considerations, and are compatible with current neural network frameworks. In that note, leading Deep Learning frameworks have already integrated support for Bayesian learning [239], which are soon bound to overtake regular Deep Learning in popularity [15].

A promising pathway to bridge explicit variance learning rules and neurobiology lies in predictive coding models. These models empirically approximate gradient backpropagation, a cornerstone technique for training deep neural networks (see chapter 3), across arbitrary computational graphs [212]. This suggests that practically any deep neural network could be trained using units dedicated to predictions and their errors. Practical applications of this theory are already evident in simpler datasets like MNIST [213]. Notably, as we detailed in the conclusion of chapter 5, the implementation of inverse variance-weighting in this context is a novel proposition, as identity matrices are often use for simplicity's sake. The programming of these networks is elegantly simple - less than 100 Python code lines for an implementation as mathematically framed here [213]. On modern hardware, they also achieve comparable accuracy and convergence speeds to traditional backpropagation techniques.

Moreover, predictive coding's use in causal inference (as a generative model) could shed new light on how neural networks simulate and analyze dynamics within their nodes. In the recent context of Large Language Models, the inherent flexibility of predictive coding models allows any unit in the network to be set as a latent variable, endowing these models with the ability for flexible internal conditioning. This would enhance steerability over the network, for desirable safety purposes. The generative nature of these models also equips them to deal with incomplete inputs or outputs, a situation which typically requires expensive data curation by humans in the context of Deep Neural Networks. This attribute could be advantageous in creating models that infer missing information seamlessly. Finally, predictive coding also extends in a rather straightforward manner to arbitrary time-space variables, which allows for dynamical modelling, something which backpropagations techniques have always struggled with [212].

In the realm of hardware, the benefits of employing variance weighting through lateral inhibition might mirror those observed in neurobiology. Specifically, this

method can offer computational allocation for significant, unpredictable fluctuations in data, while concurrently diminishing or sidestepping routine, predictable data streams. Highlighting these pronounced shifts could streamline the data transmitted across physical conduits, addressing a primary source of heat and computational inefficiency in neuromorphic chips [79, 253]. This could lead to faster, more energy-efficient neuromorphic networks that are particularly useful for edge computing, which, arguably, might also be why they are also implemented in the first place in the brain. Such asymmetrical message-passing structure is also reflected in the activity of V1, notably, in the rasterplots presented in chapter 4.

7.3. Limitations of the Studies

The elaboration of a doctoral manuscript requires the doctoral candidate to perform scientific work, but also to learn how to even do such scientific work in the first place. As such, the erudite reader might have already spotted a number of suboptimal procedures in the experiments which are intrinsically bound to the learning curve the author went through. On a positive note, none of these limitations are unsurmountable roadblocks, and most of them can serve to formulate useful hypothesis for future research directions.

First and foremost, there is an intrinsic limit in the interpretations one can give from the experiments carried in chapter 4. Such extracellular recordings were performed under anesthesia, which yields heavy modulations of the activity of neurons [331]. We have addressed this limitation in the conclusion of chapter 4's article, citing examples of experiments that were successfully translated from anesthetized to awake animals. One key neurophysiological element is however missing. Convincing experimental proofs have been published during the thesis, showing that the main effect of anesthesia is to decouple basal and apical dendrites of layer V cortical neurons [307]. Further, since pharmacological agents modulating the internal representations of the brain (i.e., predictions) have been shown to target specifically these same neurons [129], it seems evident that one part of the prediction/prediction error network was heavily modulated in our experiments. Carrying experiments in awake, vigil animals is one way to address this limit, and the experiments in chapter 5 are a fair proof that chapter 4's study can yield interesting insights into a more ecologically relevant setting.

However, these experiments were carried with Utah Arrays (see chapter 4 and 5), which means a complete loss of information of the laminar position of neurons recorded. As such, we currently have no proof of a laminar-dependent inverse variance computations in awake cortex of primates. Change of coupling in deep layers might make for a very compelling case, which could be that inverse variance is also computed for predictions in layer V. Anatomically, there is actually as much, if not more (proportionally) recurrent connectivity between neurons of layer 5 of different neighboring cortical columns [36, 20]. Functionally, layer V recurrent interactions also support very complex types of computations[328], which are unique to these layers (some similarities exist [103]). This fully-lateral model of the cortex will be the cen-

tral point of future research direction, and will be discussed in the next Perspectives section.

Second, the choice of recording local, laminar, extracellular potentials is clearly not ideal for studying recurrent interactions. Using a better suited method would have required to know beforehand that recurrent interactions can carry out inverse variance weighting in the cortex, which was only hinted at by speculative models [20, 92]. As such, a vast portion of neuronal "dark matter" was not observed, but only inferred. In a sense, this serves the interdisciplinar argument of this manuscript, which uses computational models to overcome experimental limitations. Now that these recurrent principles have been fairly well established, optical methods would be much better suited to study the inverse variance weighting mechanism in the cortex. One could expect some very interesting parallels between the results presented here, and the travelling waves (see chapter 4) observed both at the cortical area scale [51] and at the brain scale [217, 9]. In terms of dynamics, travelling waves propagate locally with speeds that are coherent with the idea of a series of iterative recurrence-based computations [51, 50]. This possesses interesting ties with predictive processing, because local recurrence can allow a visual map to predict future position of a moving object through local interactions [37], which is actually embedded within a travelling wave [26]. This experimental limitation, and its possibles interpretations in terms of travelling waves, will form part of the research perspectives discussed in the next section.

7.4. Perspectives for Future Research

To conclude this thesis, we will develop here two major research perspectives that have stemmed from the present findings, and that are actively under study at the time of this writing. The first perspective stems from chapter 4 and 5 experiments, and concern extension of this thesis' work to rodent models. The second perspective is more ambitious, and better suited for exploration during a post-doctoral tenure. Both trajectories engage with the notion that existing models of recurrent connections in the cortex may be either limited by primate-centric considerations, or could simply be fundamentally unable to account for the meso-scale implementation of message passing in the cortex.

7.4.1. Precision-Weighting without an Orientation Map

The take-home message of this manuscript is that recurrence between specifically clustered groups of cells in orientation space mediates the computation of the inverse variance of orientation in V1. But, how can these computations take place when there is no such thing as a specifically clustered groups of cells ? This is a non-trivial question, as this seemingly random architecture constitutes the organization of rodents' salt-and-pepper V1. While primates and felines have clustered orientation processors organized in a map, which emerges through their heavy reliance on vision, rodents

instead rely on whisker sensing, and thus lack a topologically structured orientation selective activity. Their somatosensory cortex, which processes information from the whiskers, do follow a topological structure. The extension of inverse variance-weighting in rodent cortex could be achieved by designing a class of stimulus similar to Motion Clouds, in the texture domain, and observe whether the variance-weighting in V1 holds in the somatosensory domain.

Back to the visual cortex, preliminary results from the co-authors Lamyae Ikan and Nelson Cortes suggest that mice do not exhibit any type of resilience to increased sensory variance, contrary to felines (chapter 4) and primates (chapter 5). If we put this under the notion that structured cortical recurrence is needed for such computations, this is one more argument in favor of this thesis' hypothesis. If we put this in the general predictive coding notion that high variance input are "explained away" and do not update the model, then it would be interesting to see what happens behaviorally. Even without processing of variance, does Bayesian inference still occur, and does the animal discard entirely its sense of vision to rely on its whiskers ?

This also speak to an interesting property of resilience to general sources of uncertainty, as opposed to solely variance. Rodents do not possess orientation maps, but do possess individual neurons strongly tuned to specific orientations, with functional properties such as contrast invariance like primates [88]. This shows that an orientation map is not crucial for feature sensitivity, nor for maintaining contrast invariance. Changes in contrast, much like changes in orientation variance, are a source of visual uncertainty. Both type of uncertainty seem to affect V1 in the same way, at least in the pionnering study of Goris et al. [115] (see also Figure 7.2. There is then a discrepancy between the notion that activity in primate V1 is similarly influenced by uncertainty of orientation (variance) and luminance (contrast), and the fact that mice are resilient to the latter but not the former. Contrast invariance mechanisms are a very popular topic in visual neuroscience, and there exists many theories on the emergence of contrast-invariant activity in V1 [330]. While some authors suppose that cortical interactions could create contrast invariance [320], others have supposed that non-linearity in the thalamo-cortical connections from the LGN is responsible for contrast invariance [88]. It would seem that, given this uncoupling between variance and luminance, the latter hypothesis would be better supported by our results.

7.4.2. The Microcircuit is Dead, Long Live the Microcircuit

The theoretical framework upon which we have built our hypotheses, predictive processing [254, 92], offers a solid mathematical basis to understand cortical functions. The accurate mapping of theory to biology, however, remains speculative [20, 283], aside from accurate understanding of the prediction error circuits [162, 163]. Recall that earlier in the text, we mentioned recent findings by Heindorf and Keller [129], who, utilizing antipsychotics and widefield calcium imaging in behaving mice, have demonstrated that antipsychotic drugs selectively impact Layer 5 neurons. As antipsychotics should target the neural elements responsible for internal representations [94] (i.e., predictions), this argues in favor of Layer 5 as the seat of predictions in the cortex.
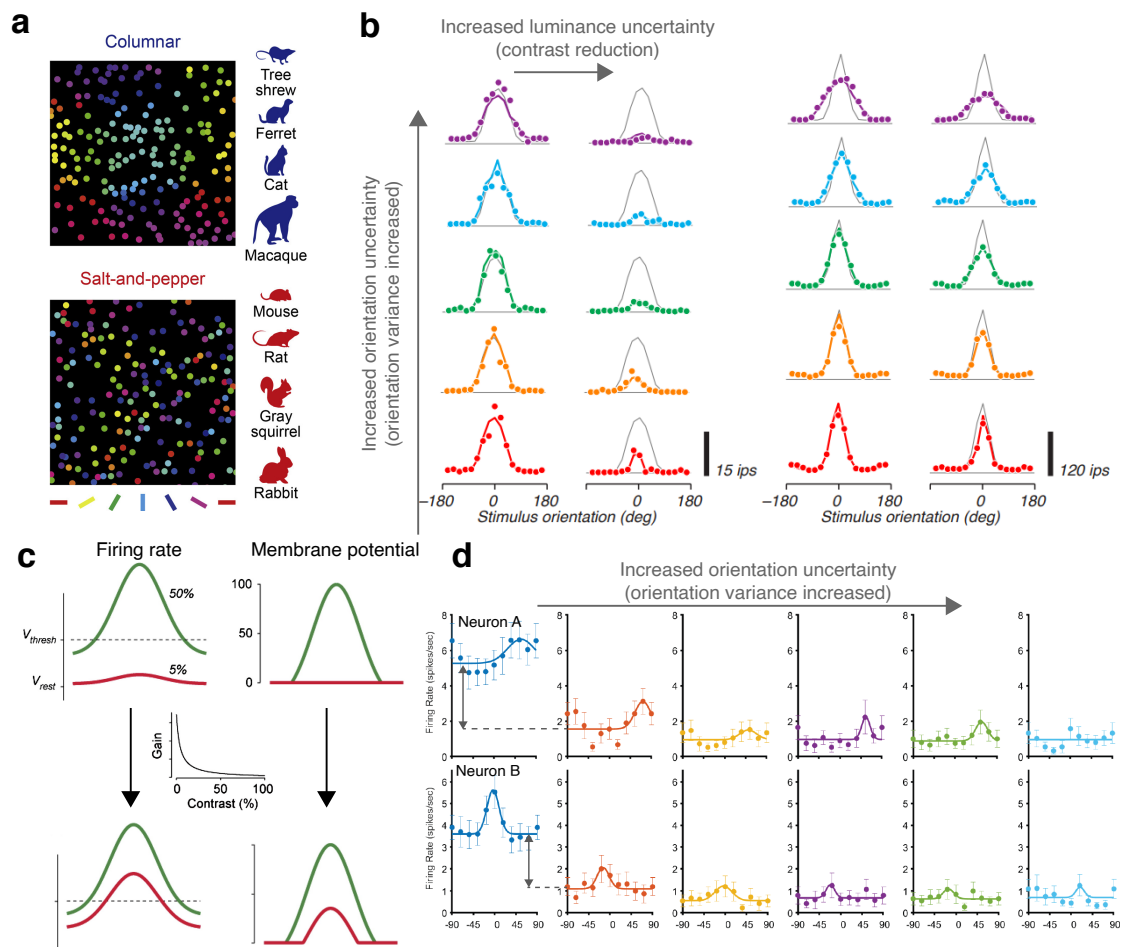
Figure 7.2. – Contrast and orientation invariance in mouse V1. (a) Salt-and-pepper or-
ganisation of V1, from [144]. (b) Similar effect of contrast and orientation
invariance on two primate V1 neurons, from [115]. (c) An example of non-
linear gain model that accounts for contrast resilience in V1, from [88].
(d) Example of mice V1 neurons, showing a massive decrease in firing rate
(arrows) with increased orientation variance (courtesy of Nelson Cortes).

This is in line with the speculative layout of predictive coding in the cortical microcircuit [20] (Figure 2.14). Yet, the effect of antipsychotics is selective to Layer 5, and leaves superficial layers unaffected. This poses a significant challenge to the established model of a series of hierarchically organized vertical microcircuits, derived from the columnar model of the cortex (Figure 7.3).

Namely, the canonical microcircuit [201] posits an input in Layer 4, followed by a processing in Layer 2/3, which is then sent to a higher order cortical area's Layer 4. This challenge is not recent, as the vertical microcircuit as always had scientific opposition [135], mostly on functional basis. Support for a new "horizontal microcircuit" perspective is now further bolstered by anatomical studies, which revealed a bidirectional flow of pathways segregated into two supra- and infragranular streams [199]. This makes intuitive evolutionary sense, given that the appearance of six cortical layers is thought to result from a duplication of a three-layered structure [278]. It thus becomes increasingly plausible that neural activity is focused on integrating disparate features to construct probabilistic representations of the environment, rather than integrating information within a vertical microcircuit. This is evidenced not only by the fact that horizontal communication is the dominant mode of the cortex [217], but also by the fact that inverse variance weighting requires a horizontal (recurrent) communication system [178].

In light of these evidences, it is essential to address the implications they hold for the cortical model of the canonical microcircuit, particularly as they propose both an extension and a challenge to the results of this thesis. If Layer 5 neurons have been affected due to anesthetic decoupling, as proposed above [307], the impact of this experimental limitation is more significant than previously thought. These neurons might not have just been part of a vertical microcircuit, but could have been an integral component of an independent horizontal deep circuit altogether, with an independent recurrent dynamic. Additionally, the manipulation of the variance of Motion Clouds and sensory inputs might have focused the present work on the inverse variance-weighting of prediction errors, rather than a combined investigation that includes both prediction errors and sensory inputs. This focus could have led to the inability to read out input variance from deep cortical layers in chapter 4. If we wish to amend this, there is an intrinsic technological challenge for current research methods of investigation (Figure 4.7).

Sampling the recurrent activity of inverse variance-weighting throughout the cortex, not at the scale of a single area, but at the dominant scale of cortex-wide activity, requires an experimental paradigm shift. This involves sampling widefield activity at the Nyquist frequency of the fastest signals required, i.e., ≈ 50Hz gamma-band prediction errors (Figure 6.2). The advent of advanced optical methods now permits kHz-level recording of brain activity using Genetically Encoded Voltage Indicators. Wide-field microscopy techniques enable the observation of awake, behaving mice, granting comprehensive access to the visual cortex's surface. By employing genetically specific constructs, it is possible to isolate recordings to deep or superficial cortical layers [194]. Through such recordings of traveling waves, we can apply the vast body of literature on brain oscillations to living neural activity.
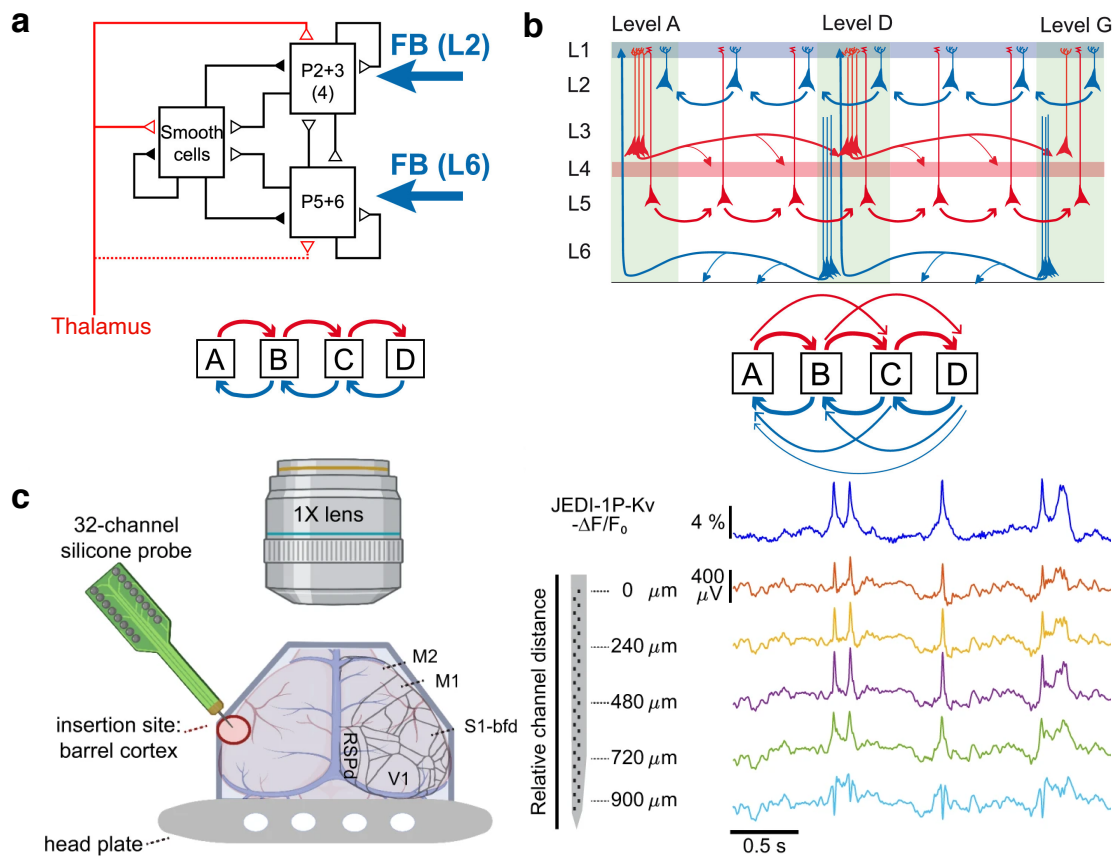
Figure 7.3. – Towards a horizontal microcircuit. (a) Canonical microcircuit of V1 [201] (top) and assumed but wrong serial communication between area (bottom). (b) The cortex instead projects in two counterstreams that densely connect every area to one another. Adapted from [199]. (c) Recording from a Genetically Encoded Voltage Indicators correlate with silicon probes, as used in this thesis, allowing for KHz-fast widefield imaging of the communication throughout the cortex, from [194].

This approach aligns with the theoretical models illustrated in Figure 6.2, facilitating manipulation of the sensorium of behaving brains to pinpoint the anatomical seat of internal neuronal representations. Correlated with the known existence of prediction error circuits, this is the next scientific step for predictive coding, which has already achieved Marr's algorithmic [92] and computational [294] levels, and is now poised to achieve implementational level, through a wholistic mapping of the neural elements of predictive coding.

Bibliography

[1]   Reza Abbas Farishta, Denis Boire, and Christian Casanova. "Hierarchical organization of corticothalamic projections to the pulvinar". In: *Cerebral Cortex Communications* 1.1 (2020), tgaa030 (cit. on p. 20).

[2]   Larry F Abbott and Frances S Chance. "Drivers and modulators from push-pull and balanced synaptic input". In: *Progress in brain research* 149 (2005), pp. 147–155 (cit. on p. 134).

[3]   Rick A Adams, Klaas Enno Stephan, Harriet R Brown, et al. "The computational anatomy of psychosis". In: *Frontiers in psychiatry* 4 (2013), p. 47 (cit. on pp. 45, 53).

[4]   Hillel Adesnik and Massimo Scanziani. "Lateral competition for cortical space by layer-specific horizontal circuits". In: *Nature* 464.7292 (2010), pp. 1155–1160 (cit. on p. 51).

[5]   Miguel Aguilera, Beren Millidge, Alexander Tschantz, et al. "How particular is the physics of the free energy principle?" In: *Physics of Life Reviews* 40 (2022), pp. 24–50 (cit. on p. 40).

[6]   Vivien Ainley, Matthew AJ Apps, Aikaterini Fotopoulou, et al. "'Bodily precision': a predictive coding account of individual differences in interoceptive accuracy". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 371.1708 (2016), p. 20160003 (cit. on p. 53).

[7]   Laurence Aitchison and Máté Lengyel. "With or without you: predictive coding and Bayesian inference in the brain". In: *Current opinion in neurobiology* 46 (2017), pp. 219–227 (cit. on p. 23).

[8]   Andrea Alamia, Dario Gordillo, Eka Chkonia, et al. "Oscillatory traveling waves reveal predictive coding abnormalities in schizophrenia". In: *bioRxiv* (2023), pp. 2023–10 (cit. on p. 164).

[9]   Andrea Alamia and Rufin VanRullen. "Alpha oscillations and traveling waves: Signatures of predictive coding?" In: *PLoS Biology* 17.10 (2019), e3000487 (cit. on pp. 165, 168).

[10]   Kimmo Alho. "Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes". In: *Ear and hearing* 16.1 (1995), pp. 38–51 (cit. on p. 52).

[11] 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". In: *Progress in brain research* 154 (2006), pp. 93–120 (cit. on pp. 28, 91, 93).

[12] Alessandra Angelucci, Jonathan B Levitt, Emma JS Walton, et al. "Circuits for local and global signal integration in primary visual cortex". In: *Journal of Neuroscience* 22.19 (2002), pp. 8633–8646 (cit. on p. 51).

[13] Michael J Arcaro, Mark A Pinsk, and Sabine Kastner. "The anatomical and functional organization of the human visual pulvinar". In: *Journal of Neuroscience* 35.27 (2015), pp. 9848–9871 (cit. on p. 133).

[14] Wyeth Bair, James R Cavanaugh, and J Anthony Movshon. "Time course and time-distance relationships for surround suppression in macaque V1 neurons". In: *Journal of Neuroscience* 23.20 (2003), pp. 7690–7701 (cit. on p. 52).

[15] Breck Baldwin. "Deep learning does not replace Bayesian modeling: Comparing research use via citation counting". In: *Applied AI Letters* 3.1 (2022), e62 (cit. on p. 166).

[16] Horace Barlow. "Redundancy reduction revisited". In: *Network: computation in neural systems* 12.3 (2001), p. 241 (cit. on p. 49).

[17] Horace B Barlow. "Retinal noise and absolute threshold". In: *Josa* 46.8 (1956), pp. 634–639 (cit. on pp. 31, 38, 56).

[18] Horace B Barlow et al. "Possible principles underlying the transformation of sensory messages". In: *Sensory communication* 1.01 (1961), pp. 217–233 (cit. on pp. 31, 57, 117, 123, 162, 163).

[19] Simon Barthelmé and Pascal Mamassian. "Evaluation of objective uncertainty in the visual system". In: *PLoS computational biology* 5.9 (2009), e1000504 (cit. on p. 36).

[20] Andre M Bastos, W Martin Usrey, Rick A Adams, et al. "Canonical microcircuits for predictive coding". In: *Neuron* 76.4 (2012), pp. 695–711 (cit. on pp. 28, 49, 50, 93, 133, 136, 137, 167–169, 171).

[21] André M Bastos, Mikael Lundqvist, Ayan S Waite, et al. "Layer and rhythm specificity for predictive routing". In: *Proceedings of the National Academy of Sciences* 117.49 (2020), pp. 31459–31469 (cit. on p. 51).

[22] Andre Moraes Bastos, Julien Vezoli, Conrado Arturo Bosman, et al. "Visual areas exert feedforward and feedback influences through distinct frequency channels". In: *Neuron* 85.2 (2015), pp. 390–401 (cit. on pp. 51, 137).

[23] Pierre Baudot, Manuel Levy, Olivier Marre, et al. "Animation of natural scene by virtual eye-movements evokes high precision and low noise in V1 neurons". In: *Frontiers in neural circuits* 7 (2013), p. 206 (cit. on p. 33).

[24]   William HA Beaudot and Kathy T Mullen. "Orientation discrimination in human vision: Psychophysics and modeling". In: *Vision Research* 46.1-2 (2006), pp. 26–46 (cit. on p. 36).

[25]   Pascal Belin, Patricia EG Bestelmeyer, Marianne Latinus, et al. "Understanding voice perception". In: *British Journal of Psychology* 102.4 (2011), pp. 711–725 (cit. on p. 87).

[26]   Gabriel B Benigno, Roberto C Budzinski, Zachary W Davis, et al. "Waves traveling over a map of visual space can ignite short-term predictions of sensory input". In: *Nature Communications* 14.1 (2023), p. 3409 (cit. on p. 168).

[27]   Philipp Berens, Alexander S Ecker, R James Cotton, et al. "A fast and simple population code for orientation in primate V1". In: *Journal of Neuroscience* 32.31 (2012), pp. 10618–10626 (cit. on p. 114).

[28]   Tom Binzegger, Rodney J Douglas, and Kevan AC Martin. "A quantitative map of the circuit of cat primary visual cortex". In: *Journal of Neuroscience* 24.39 (2004), pp. 8441–8453 (cit. on p. 28).

[29]   Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*. Vol. 4. 4. Springer, 2006 (cit. on p. 98).

[30]   Martin Boerlin, Christian K Machens, and Sophie Denève. "Predictive coding of dynamical variables in balanced spiking networks". In: *PLoS computational biology* 9.11 (2013), e1003258 (cit. on p. 51).

[31]   Rafal Bogacz. "A tutorial on the free-energy framework for modelling perception and learning". In: *Journal of mathematical psychology* 76 (2017), pp. 198–211 (cit. on pp. 38, 48, 112).

[32]   Adrian G Bondy, Ralf M Haefner, and Bruce G Cumming. "Feedback determines the structure of correlated variability in primary visual cortex". In: *Nature neuroscience* 21.4 (2018), pp. 598–606 (cit. on p. 117).

[33]   Victor Boutin, Angelo Franciosini, Frederic Chavane, et al. "Sparse deep predictive coding captures contour integration capabilities of the early visual system". In: *PLoS computational biology* 17.1 (2021), e1008629 (cit. on p. 49).

[34]   Victor Boutin, Angelo Franciosini, Frédéric Chavane, et al. "Pooling strategies in V1 can account for the functional and structural diversity across species". In: *PLOS Computational Biology* 18.7 (2022), e1010270 (cit. on pp. 29, 49, 50).

[35]   Victor Boutin, Angelo Franciosini, Frédéric Y Chavane, et al. "Sparse Deep Predictive Coding Captures Contour Integration Capabilities of the Early Visual System". In: *PLoS Computational Biology* (2020). DOI: 10.1371/journal.pcbi.1008629 (cit. on p. 61).

[36]   Farran Briggs and W Martin Usrey. "Corticogeniculate feedback and visual processing in the primate". In: *The Journal of physiology* 589.1 (2011), pp. 33–40 (cit. on pp. 27, 167).

[37]   Kenneth H Britten, Michael N Shadlen, William T Newsome, et al. "The analysis of visual motion: a comparison of neuronal and psychophysical performance". In: *Journal of Neuroscience* 12.12 (1992), pp. 4745–4765 (cit. on p. 168).

[38]   Harriet Brown, Karl Friston, and Sven Bestmann. "Active inference, attention, and motor preparation". In: *Frontiers in psychology* 2 (2011), p. 218 (cit. on p. 164).

[39]   Peter Brown. "Abnormal oscillatory synchronisation in the motor system leads to impaired movement". In: *Current opinion in neurobiology* 17.6 (2007), pp. 656–664 (cit. on p. 160).

[40]   Michael Bruyns-Haylett, Jingjing Luo, Aneurin J Kennerley, et al. "The neurogenesis of P1 and N1: A concurrent EEG/LFP study". In: *Neuroimage* 146 (2017), pp. 575–588 (cit. on p. 136).

[41]   Stephane Bugeon, Joshua Duffield, Mario Dipoppa, et al. "A transcriptomic axis predicts state modulation of cortical interneurons". In: *Nature* 607.7918 (2022), pp. 330–338 (cit. on pp. 50, 51).

[42]   Ed Bullmore and Olaf Sporns. "The economy of brain network organization". In: *Nature reviews neuroscience* 13.5 (2012), pp. 336–349 (cit. on p. 120).

[43]   György Buzsáki, Costas A Anastassiou, and Christof Koch. "The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes". In: *Nature reviews neuroscience* 13.6 (2012), pp. 407–420 (cit. on p. 95).

[44]   Jérémy Camon, Sandrine Hugues, Melissa A Erlandson, et al. "The timing of sensory-guided behavioral response is represented in the mouse primary somatosensory cortex". In: *Cerebral Cortex* 29.7 (2019), pp. 3034–3047 (cit. on p. 87).

[45]   Matteo Carandini and David J Heeger. "Normalization as a canonical neural computation". In: *Nature Reviews Neuroscience* 13.1 (2012), pp. 51–62 (cit. on p. 87).

[46]   C Casanova. "Functions of the pulvinar in vision". In: *The Visual Neurosciences* (2004), pp. 592–608 (cit. on pp. 28, 132, 164).

[47]   C Casanova, L Merabet, A Desautels, et al. "Higher-order motion processing in the pulvinar". In: *Progress in brain research* 134 (2001), pp. 71–82 (cit. on p. 53).

[48]   Frances S Chance, Sacha B Nelson, and Larry F Abbott. "Complex cells as cortically amplified simple cells". In: *Nature neuroscience* 2.3 (1999), pp. 277–282 (cit. on pp. 29, 91, 112).

[49]   Barbara Chapman, Kathleen R Zahs, and Michael P Stryker. "Relation of cortical cell orientation selectivity to alignment of receptive fields of the geniculocortical afferents that arborize within a single orientation column in ferret visual cortex". In: *Journal of Neuroscience* 11.5 (1991), pp. 1347–1358 (cit. on p. 90).

177

[50]   Frédéric Chavane, Laurent Udo Perrinet, and James Rankin. "Revisiting horizontal connectivity rules in V1: from like-to-like towards like-to-all". In: *Brain Structure and Function* 227.4 (2022), pp. 1279–1295 (cit. on pp. 50, 51, 90, 168).

[51]   Frédéric Chavane, Dahlia Sharon, Dirk Jancke, et al. "Lateral spread of orientation selectivity in V1 is controlled by intracortical cooperativity". In: *Frontiers in systems neuroscience* 5 (2011), p. 4 (cit. on pp. 28, 52, 55, 90, 168).

[52]   S Chemla and F Chavane. "Voltage-sensitive dye imaging: Technique review and models". In: *Journal of Physiology-Paris* 104.1-2 (2010), pp. 40–50 (cit. on pp. 94, 96).

[53]   Sandrine Chemla and Frederic Chavane. "A biophysical cortical column model to study the multi-component origin of the VSDI signal". In: *NeuroImage* 53.2 (2010), pp. 420–438 (cit. on p. 96).

[54]   Xiaowei Chen, Ulrich Leischner, Nathalie L Rochefort, et al. "Functional mapping of single spines in cortical neurons in vivo". In: *Nature* 475.7357 (2011), pp. 501–505 (cit. on p. 55).

[55]   Xing Chen, Aitor Morales-Gregorio, Julia Sprenger, et al. "1024-channel electrophysiological recordings in macaque V1 and V4 during resting state". In: *Scientific data* 9.1 (2022), p. 77 (cit. on p. 118).

[56]   Adrian Cheng, J Tiago Goncalves, Peyman Golshani, et al. "Simultaneous two-photon calcium imaging at different depths with spatiotemporal multiplexing". In: *Nature methods* 8.2 (2011), pp. 139–142 (cit. on p. 96).

[57]   SueYeon Chung and LF Abbott. "Neural population geometry: An approach for understanding biological and artificial neural networks". In: *Current opinion in neurobiology* 70 (2021), pp. 137–144 (cit. on p. 114).

[58]   Oliver M Cliff, Annie G Bryant, Joseph T Lizier, et al. "Unifying pairwise interactions in complex dynamics". In: *Nature Computational Science* (2023), pp. 1–11 (cit. on p. 118).

[59]   Matteo Colombo and Peggy Seriès. "Bayes in the brain—on Bayesian modelling in neuroscience". In: *The British journal for the philosophy of science* (2012) (cit. on p. 23).

[60]   David M Coppola, Harriett R Purves, Allison N McCoy, et al. "The distribution of oriented contours in the real world". In: *Proceedings of the National Academy of Sciences* 95.7 (1998), pp. 4002–4006 (cit. on p. 58).

[61]   Nelson Cortes, Reza Abbas Farishta, Hugo J Ladret, et al. "Corticothalamic Projections Gate Alpha Rhythms in the Pulvinar". In: *Frontiers in Cellular Neuroscience* 15 (2021), p. 787170 (cit. on pp. 51, 134, 137).

[62]   Nelson Cortes, Hugo J Ladret, Reza Abbas-Farishta, et al. "The pulvinar as a hub of visual processing and cortical integration". In: *Trends in Neurosciences* (2023) (cit. on pp. 135, 208).

[63] W Noel Cottingham and Derek A Greenwood. *An introduction to the standard model of particle physics*. Cambridge university press, 2023 (cit. on p. 22).

[64] Robert Coultrip, Richard Granger, and Gary Lynch. "A cortical model of winner-take-all competition via lateral inhibition". In: *Neural networks* 5.1 (1992), pp. 47–54 (cit. on p. 50).

[65] Elise N Covic and S Murray Sherman. "Synaptic properties of connections between the primary and secondary auditory cortices in mice". In: *Cerebral Cortex* 21.11 (2011), pp. 2425–2441 (cit. on p. 93).

[66] OD Creutzfeldt. "Extrageniculo-striate visual mechanisms: compartmentalization of visual functions". In: *Progress in Brain Research* 75 (1988), pp. 307–320 (cit. on p. 134).

[67] Laszlo Csanky. "Fast parallel matrix inversion algorithms". In: *16th Annual Symposium on Foundations of Computer Science (sfcs 1975)*. IEEE. 1975, pp. 11–12 (cit. on p. 48).

[68] Jozsef Csicsvari, Darrell A Henze, Brian Jamieson, et al. "Massively parallel recording of unit and local field potentials with silicon-based electrodes". In: *Journal of neurophysiology* 90.2 (2003), pp. 1314–1323 (cit. on p. 95).

[69] Jean-Bernard Damasse, Laurent U Perrinet, Laurent Madelain, et al. "Reinforcement effects in anticipatory smooth eye movements". In: *Journal of Vision* 18.11 (2018), pp. 14–14 (cit. on p. 112).

[70] Mircea V Diudea and Ivan Gutman. "Wiener-type topological indices". In: *Croatica chemica acta* 71.1 (1998), pp. 21–51 (cit. on p. 123).

[71] Karl-Anton Dorph-Petersen and David A Lewis. "Postmortem structural studies of the thalamus in schizophrenia". In: *Schizophrenia Research* 180 (2017), pp. 28–35 (cit. on p. 163).

[72] Zongker Doug. "Chicken chicken chicken: Chicken chicken". In: *Annals of Improbable Research* 12.5 (2006), pp. 16–21. URL: http://improbable.com/airchives/paperair/volume12/v12i5/chicken-12-5.pdf (cit. on p. 44).

[73] Rodney J Douglas and Kevan AC Martin. "Neuronal circuits of the neocortex". In: *Annu. Rev. Neurosci.* 27 (2004), pp. 419–451 (cit. on pp. 28, 50).

[74] Rodney J Douglas, Kevan AC Martin, and David Whitteridge. "A canonical microcircuit for neocortex". In: *Neural computation* 1.4 (1989), pp. 480–488 (cit. on pp. 28, 50, 55, 89).

[75] Laura Dugué, Philippe Marque, and Rufin VanRullen. "The phase of ongoing oscillations mediates the causal relation between brain excitation and visual perception". In: *Journal of neuroscience* 31.33 (2011), pp. 11889–11893 (cit. on p. 165).

[76] Andreas K Engel and Pascal Fries. "Beta-band oscillations—signalling the status quo?" In: *Current opinion in neurobiology* 20.2 (2010), pp. 156–165 (cit. on p. 51).

[77]   Andreas K Engel, Pascal Fries, and Wolf Singer. "Dynamic predictions: oscillations and synchrony in top–down processing". In: *Nature Reviews Neuroscience* 2.10 (2001), pp. 704–716 (cit. on p. 51).

[78]   Andreas K Engel, Peter König, Andreas K Kreiter, et al. "Temporal coding in the visual cortex: new vistas on integration in the nervous system". In: *Trends in neurosciences* 15.6 (1992), pp. 218–226 (cit. on p. 97).

[79]   Jason K Eshraghian, Xinxin Wang, and Wei D Lu. "Memristor-based binarized spiking neural networks: Challenges and applications". In: *IEEE Nanotechnology Magazine* 16.2 (2022), pp. 14–23 (cit. on pp. 85, 167).

[80]   A Aldo Faisal, Luc PJ Selen, and Daniel M Wolpert. "Noise in the nervous system". In: *Nature reviews neuroscience* 9.4 (2008), pp. 292–303 (cit. on p. 45).

[81]   A Aldo Faisal, John A White, and Simon B Laughlin. "Ion-channel noise places limits on the miniaturization of the brain's wiring". In: *Current Biology* 15.12 (2005), pp. 1143–1149 (cit. on p. 45).

[82]   Farzad V Farahani, Waldemar Karwowski, and Nichole R Lighthall. "Application of graph theory for identifying connectivity patterns in human brain networks: a systematic review". In: *frontiers in Neuroscience* 13 (2019), p. 585 (cit. on p. 119).

[83]   Daniel J Felleman and David C Van Essen. "Distributed hierarchical processing in the primate cerebral cortex." In: *Cerebral cortex (New York, NY: 1991)* 1.1 (1991), pp. 1–47 (cit. on pp. 50, 52).

[84]   Enrique J Fernández, Ignacio Iglesias, and Pablo Artal. "Closed-loop adaptive optics in the human eye". In: *Optics letters* 26.10 (2001), pp. 746–748 (cit. on p. 26).

[85]   David Ferster. "Orientation selectivity of synaptic potentials in neurons of cat primary visual cortex". In: *Journal of Neuroscience* 6.5 (1986), pp. 1284–1301 (cit. on p. 90).

[86]   David Ferster, Sooyoung Chung, and Heidi Wheat. "Orientation selectivity of thalamic input to simple cells of cat visual cortex". In: *Nature* 380.6571 (1996), pp. 249–252 (cit. on p. 90).

[87]   David J Field. "Relations between the statistics of natural images and the response properties of cortical cells". In: *Josa a* 4.12 (1987), pp. 2379–2394 (cit. on pp. 33, 35).

[88]   Ian M Finn, Nicholas J Priebe, and David Ferster. "The emergence of contrast-invariant orientation tuning in simple cells of cat visual cortex". In: *Neuron* 54.1 (2007), pp. 137–152 (cit. on pp. 169, 170).

[89]   Christopher D Fiorillo, William T Newsome, and Wolfram Schultz. "The temporal precision of reward prediction in dopamine neurons". In: *Nature neuroscience* 11.8 (2008), pp. 966–973 (cit. on p. 164).

[90]   József Fiser, Chiayu Chiu, and Michael Weliky. "Small modulation of ongoing cortical dynamics by sensory input during natural vision". In: *Nature* 431.7008 (2004), pp. 573–578 (cit. on pp. 33, 52).

[91]   Johannes Friedrich, Pengcheng Zhou, and Liam Paninski. "Fast online deconvolution of calcium imaging data". In: *PLoS computational biology* 13.3 (2017), e1005423 (cit. on p. 96).

[92]   Karl Friston. "A theory of cortical responses". In: *Philosophical transactions of the Royal Society B: Biological sciences* 360.1456 (2005), pp. 815–836 (cit. on pp. 40, 43, 45, 48, 52, 53, 168, 169, 173).

[93]   Karl Friston. "Predictive coding, precision and synchrony". In: *Cognitive neuroscience* 3.3-4 (2012), pp. 238–239 (cit. on p. 53).

[94]   Karl Friston. "Does predictive coding have a future?" In: *Nature neuroscience* 21.8 (2018), pp. 1019–1021 (cit. on pp. 49, 169).

[95]   Karl Friston. "Computational psychiatry: from synapses to sentience". In: *Molecular psychiatry* 28.1 (2023), pp. 256–268 (cit. on p. 163).

[96]   Karl Friston, Rick A Adams, Laurent Perrinet, et al. "Perceptions as hypotheses: saccades as experiments". In: *Frontiers in psychology* 3 (2012), p. 151 (cit. on p. 26).

[97]   Karl Friston, Thomas FitzGerald, Francesco Rigoli, et al. "Active inference and learning". In: *Neuroscience & Biobehavioral Reviews* 68 (2016), pp. 862–879 (cit. on p. 89).

[98]   Karl Friston and Stefan Kiebel. "Predictive coding under the free-energy principle". In: *Philosophical transactions of the Royal Society B: Biological sciences* 364.1521 (2009), pp. 1211–1221 (cit. on pp. 40, 49).

[99]   Karl Friston, James Kilner, and Lee Harrison. "A free energy principle for the brain". In: *Journal of physiology-Paris* 100.1-3 (2006), pp. 70–87 (cit. on pp. 40, 52).

[100]   Karl J Friston. "Waves of prediction". In: *PLoS biology* 17.10 (2019), e3000426 (cit. on p. 136).

[101]   Thomas MJ Fruchterman and Edward M Reingold. "Graph drawing by force-directed placement". In: *Software: Practice and experience* 21.11 (1991), pp. 1129–1164 (cit. on p. 127).

[102]   Christopher S Furmanski and Stephen A Engel. "An oblique effect in human primary visual cortex". In: *Nature neuroscience* 3.6 (2000), pp. 535–536 (cit. on p. 54).

[103]   Alessandro R Galloni, Aeron Laffere, and Ede Rancz. "Apical length governs computational diversity of layer 5 pyramidal neurons". In: *Elife* 9 (2020), e55761 (cit. on p. 167).

[104] Marta I Garrido, James M Kilner, Stefan J Kiebel, et al. "Evoked brain responses are generated by feedback loops". In: *Proceedings of the National Academy of Sciences* 104.52 (2007), pp. 20961–20966 (cit. on p. 52).

[105] Juan Manuel Garrido Wainer. "Understanding the development and use of tools in neuroscience: the case of the tungsten micro-electrode". In: *Synthese* 200.6 (2022), p. 446 (cit. on p. 95).

[106] Nicolas Gauvrit, Fernando Soler-Toscano, and Hector Zenil. "Natural scene statistics mediate the perception of image complexity". In: *Visual Cognition* 22.8 (2014), pp. 1084–1091 (cit. on p. 33).

[107] Karl R Gegenfurtner and Daniel C Kiper. "Color vision". In: *Annual review of neuroscience* 26.1 (2003), pp. 181–206 (cit. on p. 26).

[108] Wilson S Geisler, Duane G Albrecht, Alison M Crane, et al. "Motion direction signals in the primary visual cortex of cat and monkey". In: *Visual neuroscience* 18.4 (2001), pp. 501–516 (cit. on p. 54).

[109] Laura S Geurts, Ruben S van Bergen, James RH Cooke, et al. "Reported confidence is based on imprecision in visual cortical stimulus representations". In: *Journal of Vision* 21.9 (2021), pp. 1918–1918 (cit. on p. 36).

[110] Laura S Geurts, James RH Cooke, Ruben S van Bergen, et al. "Subjective confidence reflects representation of Bayesian probability in cortex". In: *Nature Human Behaviour* 6.2 (2022), pp. 294–305 (cit. on p. 36).

[111] Masoud Ghodrati, Seyed-Mahdi Khaligh-Razavi, and Sidney R Lehky. "Towards building a more complex view of the lateral geniculate nucleus: Recent advances in understanding its role". In: *Progress in Neurobiology* 156 (2017), pp. 214–255 (cit. on p. 27).

[112] Krishna K Ghosh, Sascha Bujan, Silke Haverkamp, et al. "Types of bipolar cells in the mouse retina". In: *Journal of Comparative Neurology* 469.1 (2004), pp. 70–82 (cit. on p. 26).

[113] Joshua I Glaser, Ari S Benjamin, Raeed H Chowdhury, et al. "Machine learning for neural decoding". In: *Eneuro* 7.4 (2020) (cit. on pp. 97, 114).

[114] Mitchell Glickstein and Giacomo Rizzolatti. "Francesco Gennari and the structure of the cerebral cortex". In: *Trends in Neurosciences* 7.12 (1984), pp. 464–467 (cit. on p. 28).

[115] Robbe LT Goris, Eero P Simoncelli, and J Anthony Movshon. "Origin and function of tuning diversity in macaque visual cortex". In: *Neuron* 88.4 (2015), pp. 819–831 (cit. on pp. 54, 55, 112, 169, 170).

[116] Nathan W Gouwens, Staci A Sorensen, Jim Berg, et al. "Classification of electrophysiological and morphological neuron types in the mouse visual cortex". In: *Nature neuroscience* 22.7 (2019), pp. 1182–1195 (cit. on p. 50).

[117]   Ulf Grenander and Michael I Miller. "Computational anatomy: An emerging discipline". In: *Quarterly of applied mathematics* 56.4 (1998), pp. 617–694 (cit. on p. 53).

[118]   Antoine Grimaldi, Amélie Gruel, Camille Besnainou, et al. "Precise Spiking Motifs in Neurobiological and Neuromorphic Data". In: *Brain Sciences* 13.1 (2022), p. 68 (cit. on p. 97).

[119]   Antoine Grimaldi and Laurent U Perrinet. "Learning heterogeneous delays in a layer of spiking neurons for fast motion detection". In: *Biological Cybernetics* (2023), pp. 1–15 (cit. on p. 85).

[120]   Amiram Grinvald, Edmund Lieke, Ron D Frostig, et al. "Functional architecture of cortex revealed by optical imaging of intrinsic signals". In: *Nature* 324.6095 (1986), pp. 361–364 (cit. on pp. 29, 90).

[121]   RW Guillery and SM Sherman. "The thalamus as a monitor of motor outputs". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 357.1428 (2002), pp. 1809–1821 (cit. on p. 134).

[122]   Torkel Hafting, Marianne Fyhn, Sturla Molden, et al. "Microstructure of a spatial map in the entorhinal cortex". In: *Nature* 436.7052 (2005), pp. 801–806 (cit. on p. 56).

[123]   David Hansel and Carl van Vreeswijk. "The mechanism of orientation selectivity in primary visual cortex without a functional map". In: *Journal of Neuroscience* 32.12 (2012), pp. 4049–4064 (cit. on pp. 90, 91).

[124]   Bruce C Hansen and Edward A Essock. "A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes". In: *Journal of vision* 4.12 (2004), pp. 5–5 (cit. on pp. 33, 58).

[125]   Peter Hansen, Morten Kringelbach, and Riitta Salmelin. *MEG: An introduction to methods*. Oxford university press, 2010 (cit. on p. 96).

[126]   DO Hebb. "The organization of behavior. A neuropsychological theory". In: (1949) (cit. on p. 44).

[127]   David W Heeley and Hannah M Buchanan-Smith. "The influence of stimulus shape on orientation acuity". In: *Experimental Brain Research* 120.2 (1998), pp. 217–222 (cit. on p. 36).

[128]   DW Heeley, B Timney, IR Paterson, et al. "Width discrimination for band-pass stimuli". In: *Vision Research* 29.7 (1989), pp. 901–905 (cit. on pp. 36, 54).

[129]   Matthias Heindorf and Georg B Keller. "Reduction of Layer 5 Mediated Long-Range Cortical Communication by Antipsychotic Drugs". In: *bioRxiv* (2022) (cit. on pp. 167, 169).

[130]   Hermann Ludwig Ferdinand von Helmholtz, James Powell Cocke Southall, et al. "Treatise on physiological optics". In: *(No Title)* (1925) (cit. on pp. 35, 40).

[131]   Oscar Herreras. "Local field potentials: myths and misunderstandings". In: *Frontiers in neural circuits* 10 (2016), p. 101 (cit. on p. 95).

[132]   Loreen Hertäg and Claudia Clopath. "Prediction-error neurons in circuits with multiple neuron types: Formation, refinement, and functional implications". In: *Proceedings of the National Academy of Sciences* 119.13 (2022), e2115699119 (cit. on p. 51).

[133]   David Holmes. "Reconstructing the retina." In: *Nature* 561.7721 (2018), S2–S3 (cit. on p. 26).

[134]   Guillermo Horga, Kelly C Schatz, Anissa Abi-Dargham, et al. "Deficits in predictive coding underlie hallucinations in schizophrenia". In: *Journal of Neuroscience* 34.24 (2014), pp. 8072–8082 (cit. on p. 53).

[135]   Jonathan C Horton and Daniel L Adams. "The cortical column: a structure without a function". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 360.1456 (2005), pp. 837–862 (cit. on pp. 28, 50, 171).

[136]   Gábor Horváth, Tamás Szörényi, Ádám Pereszlényi, et al. "Why do horseflies need polarization vision for host detection? Polarization helps tabanid flies to select sunlit dark host animals from the dark patches of the visual environment". In: *Royal Society Open Science* 4.11 (2017), p. 170735 (cit. on p. 24).

[137]   David H Hubel and Torsten N Wiesel. "Receptive fields of single neurones in the cat's striate cortex". In: *The Journal of physiology* 148.3 (1959), p. 574 (cit. on pp. 29, 56, 95, 97, 163).

[138]   David H Hubel and Torsten N Wiesel. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". In: *The Journal of physiology* 160.1 (1962), p. 106 (cit. on pp. 29, 30, 58, 90, 97, 112).

[139]   Bob Hutchins and BV Updyke. "Retinotopic organization within the lateral posterior complex of the cat". In: *Journal of Comparative Neurology* 285.3 (1989), pp. 350–398 (cit. on p. 133).

[140]   M Florencia Iacaruso, Ioana T Gasler, and Sonja B Hofer. "Synaptic organization of visual space in primary visual cortex". In: *Nature* 547.7664 (2017), pp. 449–452 (cit. on p. 55).

[141]   Michael Ibbotson and Bart Krekelberg. "Visual perception and saccadic eye movements". In: *Current opinion in neurobiology* 21.4 (2011), pp. 553–558 (cit. on p. 26).

[142]   Thomas R Insel. "Rethinking schizophrenia". In: *Nature* 468.7321 (2010), pp. 187–193 (cit. on p. 163).

[143]   Naoum P Issa, Christopher Trepel, and Michael P Stryker. "Spatial frequency maps in cat visual cortex". In: *Journal of Neuroscience* 20.22 (2000), pp. 8504–8514 (cit. on p. 29).

[144]   Jaeson Jang, Min Song, and Se-Bum Paik. "Retino-cortical mapping ratio predicts columnar and salt-and-pepper organization in mammalian visual cortex". In: *Cell reports* 30.10 (2020), pp. 3270–3279 (cit. on pp. 91, 170).

[145]   Jean-Nicolas Jérémie, Emmanuel Daucé, and Laurent U Perrinet. "Retinotopy improves the categorisation and localisation of visual objects in CNNs". In: *International Conference on Artificial Neural Networks*. Springer. 2023, pp. 574–584 (cit. on p. 30).

[146]   Hongbo Jia, Nathalie L Rochefort, Xiaowei Chen, et al. "Dendritic organization of sensory input to cortical neurons in vivo". In: *Nature* 464.7293 (2010), pp. 1307–1312 (cit. on p. 55).

[147]   Aaron P Johnson and Curtis L Baker. "First-and second-order information in natural images: a filter-based approach to image statistics". In: *JOSA A* 21.6 (2004), pp. 913–925 (cit. on p. 35).

[148]   Edward G Jones. *The thalamus*. Springer Science & Business Media, 2012 (cit. on p. 133).

[149]   Bela Julesz. "Textons, the elements of texture perception, and their interactions". In: *Nature* 290.5802 (1981), pp. 91–97 (cit. on p. 36).

[150]   Bela Julesz and James R Bergen. "Human factors and behavioral science: Textons, the fundamental elements in preattentive vision and perception of textures". In: *Bell system technical journal* 62.6 (1983), pp. 1619–1645 (cit. on p. 36).

[151]   Jon H Kaas, Hui-Xin Qi, and Iwona Stepniewska. "Escaping the nocturnal bottleneck, and the evolution of the dorsal and ventral streams of visual processing in primates". In: *Philosophical Transactions of the Royal Society B* 377.1844 (2022), p. 20210293 (cit. on p. 31).

[152]   Yoshinao Kajikawa and Charles E Schroeder. "How local is the local field potential?" In: *Neuron* 72.5 (2011), pp. 847–858 (cit. on p. 95).

[153]   Ryota Kanai, Yutaka Komura, Stewart Shipp, et al. "Cerebral hierarchies: predictive processing, precision and the pulvinar". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1668 (2015), p. 20140169 (cit. on pp. 45, 48, 53, 133, 135).

[154]   Eric R Kandel, James H Schwartz, Thomas M Jessell, et al. *Principles of neural science*. Vol. 4. McGraw-hill New York, 2000 (cit. on pp. 24, 25, 94).

[155]   Jun Il Kang, Frédéric Huppé-Gourgues, and Elvire Vaucher. "Boosting visual cortex function and plasticity with acetylcholine to enhance visual perception". In: *Frontiers in systems neuroscience* (2014), p. 172 (cit. on p. 164).

[156]   Prakash Kara, John S Pezaris, Sergey Yurgenson, et al. "The spatial receptive field of thalamic inputs to single cortical simple cells revealed by the interaction of visual and electrical stimulation". In: *Proceedings of the National Academy of Sciences* 99.25 (2002), pp. 16261–16266 (cit. on p. 90).

[157]   Friston Karl. "A free energy principle for biological systems". In: *Entropy* 14.11 (2012), pp. 2100–2121 (cit. on p. 40).

[158] Matthias Kaschube. "Neural maps versus salt-and-pepper organization in visual cortex". In: *Current opinion in neurobiology* 24 (2014), pp. 95–102 (cit. on p. 31).

[159] Sabine Kastner and Leslie G Ungerleider. "The neural basis of biased competition in human visual cortex". In: *Neuropsychologia* 39.12 (2001), pp. 1263–1276 (cit. on p. 29).

[160] Dennis Kätzel, Boris V Zemelman, Christina Buetfering, et al. "The columnar and laminar organization of inhibitory connections to neocortical excitatory cells". In: *Nature neuroscience* 14.1 (2011), pp. 100–107 (cit. on p. 52).

[161] DRT Keeble, FAA Kingdom, B Moulden, et al. "Detection of orientationally multimodal textures". In: *Vision Research* 35.14 (1995), pp. 1991–2005 (cit. on p. 36).

[162] Georg B Keller, Tobias Bonhoeffer, and Mark Hübener. "Sensorimotor mismatch signals in primary visual cortex of the behaving mouse". In: *Neuron* 74.5 (2012), pp. 809–815 (cit. on pp. 112, 160, 169).

[163] Georg B Keller and Thomas D Mrsic-Flogel. "Predictive processing: a canonical cortical computation". In: *Neuron* 100.2 (2018), pp. 424–435 (cit. on pp. 56, 87–89, 112, 162, 169).

[164] Alex Kendall and Yarin Gal. "What uncertainties do we need in bayesian deep learning for computer vision?" In: *Advances in neural information processing systems* 30 (2017) (cit. on pp. 165, 166).

[165] Jean-Rémi King and Stanislas Dehaene. "Characterizing the dynamics of mental representations: the temporal generalization method". In: *Trends in cognitive sciences* 18.4 (2014), pp. 203–210 (cit. on pp. 113, 114).

[166] Marcel Kinsbourne. "Somatic twist: A model for the evolution of decussation." In: *Neuropsychology* 27.5 (2013), p. 511 (cit. on p. 27).

[167] Holle Kirchner and Simon J Thorpe. "Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited". In: *Vision research* 46.11 (2006), pp. 1762–1776 (cit. on p. 31).

[168] Timo Kirschstein and Rüdiger Köhling. "What is the source of the EEG?" In: *Clinical EEG and neuroscience* 40.3 (2009), pp. 146–149 (cit. on p. 96).

[169] Ádám Koblinger, József Fiser, and Máté Lengyel. "Representations of uncertainty: where art thou?" In: *Current Opinion in Behavioral Sciences* 38 (2021), pp. 150–162 (cit. on p. 36).

[170] Peter Kok, Dobromir Rahnev, Janneke FM Jehee, et al. "Attention reverses the effect of prediction in silencing sensory signals". In: *Cerebral cortex* 22.9 (2012), pp. 2197–2206 (cit. on pp. 52, 137).

[171] Helga Kolb and Ralph Nelson. "Neural architecture of the cat retina". In: *Progress in retinal research* 3 (1984), pp. 21–60 (cit. on p. 30).

[172] Gabriel Kreiman and Thomas Serre. "Beyond the feedforward sweep: feedback computations in the visual cortex". In: *Annals of the New York Academy of Sciences* 1464.1 (2020), pp. 222–241 (cit. on p. 88).

[173] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems* 25 (2012) (cit. on p. 61).

[174] Ronald HH Kröger and Oliver Biehlmaier. "Space-saving advantage of an inverted retina". In: *Vision research* 49.18 (2009), pp. 2318–2321 (cit. on p. 26).

[175] Sandra J Kuhlman, Nicholas D Olivas, Elaine Tring, et al. "A disinhibitory microcircuit initiates critical-period plasticity in the visual cortex". In: *Nature* 501.7468 (2013), pp. 543–546 (cit. on p. 51).

[176] Hugo Ladret. "HD natural images database for sparse coding". In: *FigShare* (2023). DOI: "10.6084/m9.figshare.24167265.v1" (cit. on p. 86).

[177] Hugo J Ladret, Christian Casanova, and Laurent Udo Perrinet. "Kernel Heterogeneity Improves Sparseness of Natural Images Representations". In: *arXiv preprint arXiv:2312.14685* (2023) (cit. on p. 62).

[178] Hugo J Ladret, Nelson Cortes, Lamyae Ikan, et al. "Cortical recurrence supports resilience to sensory variance in the primary visual cortex". In: *Communications Biology* 6.1 (2023), p. 667 (cit. on pp. 32, 53–55, 96, 98, 133, 171, 224, 232).

[179] Victor AF Lamme, Hans Super, and Henk Spekreijse. "Feedforward, horizontal, and feedback processing in the visual cortex". In: *Current opinion in neurobiology* 8.4 (1998), pp. 529–535 (cit. on p. 88).

[180] Michael F Land. *Eyes to see: the astonishing variety of vision in nature*. Oxford University Press, USA, 2018 (cit. on p. 24).

[181] Richard D Lange, Camille Gómez-Laberge, Vladimir K Berezovskii, et al. "Weak evidence for neural correlates of task-switching in macaque V1". In: *Journal of Neurophysiology* 129.5 (2023), pp. 1021–1044 (cit. on pp. 117, 123).

[182] Matz Larsson. "The optic chiasm: a turning point in the evolution of eye/hand coordination". In: *Frontiers in Zoology* 10.1 (2013), pp. 1–12 (cit. on p. 27).

[183] Simon Laughlin. "A simple coding procedure enhances a neuron's information capacity". In: *Zeitschrift für Naturforschung c* 36.9-10 (1981), pp. 910–912 (cit. on p. 31).

[184] Simon B Laughlin, Rob R de Ruyter van Steveninck, and John C Anderson. "The metabolic cost of neural information". In: *Nature neuroscience* 1.1 (1998), pp. 36–41 (cit. on p. 31).

[185] Yann LeCun, Bernhard Boser, John S Denker, et al. "Backpropagation applied to handwritten zip code recognition". In: *Neural computation* 1.4 (1989), pp. 541–551 (cit. on p. 47).

[186] Yann LeCun, Léon Bottou, Yoshua Bengio, et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324 (cit. on p. 61).

[187] Olivier Ledoit and Michael Wolf. "Honey, I shrunk the sample covariance matrix". In: *UPF economics and business working paper* 691 (2003) (cit. on p. 119).

[188] Olivier Ledoit and Michael Wolf. "A well-conditioned estimator for large-dimensional covariance matrices". In: *Journal of multivariate analysis* 88.2 (2004), pp. 365–411 (cit. on p. 119).

[189] Paula Sanz Leon, Ivo Vanzetta, Guillaume S Masson, et al. "Motion clouds: model-based stimulus synthesis of natural-like random textures for the study of motion perception". In: *Journal of neurophysiology* 107.11 (2012), pp. 3217–3226 (cit. on pp. 20, 34, 62, 163).

[190] Baowang Li, Matthew R Peterson, and Ralph D Freeman. "Oblique effect: a neural basis in the visual cortex". In: *Journal of neurophysiology* 90.1 (2003), pp. 204–217 (cit. on p. 54).

[191] Xuhong Liao, Athanasios V Vasilakos, and Yong He. "Small-world human brain networks: perspectives and challenges". In: *Neuroscience & Biobehavioral Reviews* 77 (2017), pp. 286–300 (cit. on pp. 122, 130).

[192] Belle Liu, Arthur Hong, Fred Rieke, et al. "Predictive encoding of motion begins in the primate retina". In: *Nature neuroscience* 24.9 (2021), pp. 1280–1291 (cit. on p. 27).

[193] Nikos K Logothetis. "What we can do and what we cannot do with fMRI". In: *Nature* 453.7197 (2008), pp. 869–878 (cit. on p. 96).

[194] Xiaoyu Lu, Yunmiao Wang, Zhuohe Liu, et al. "Widefield imaging of rapid pan-cortical voltage dynamics with an indicator evolved for one-photon microscopy". In: *Nature Communications* 14.1 (2023), p. 6423 (cit. on pp. 171, 172).

[195] JS Lund, GH Henry, CL MacQueen, et al. "Anatomical organization of the primary visual cortex (area 17) of the cat. A comparison with area 17 of the macaque monkey". In: *Journal of Comparative Neurology* 184.4 (1979), pp. 599–618 (cit. on p. 31).

[196] John Makhoul. "Linear prediction: A tutorial review". In: *Proceedings of the IEEE* 63.4 (1975), pp. 561–580 (cit. on p. 49).

[197] Mohammad Ali Mansournia and Douglas G Altman. "Inverse probability weighting". In: *Bmj* 352 (2016) (cit. on p. 41).

[198] Nikola T Markov, P Misery, Arnaud Falchier, et al. "Weight consistency specifies regularities of macaque cortical networks". In: *Cerebral cortex* 21.6 (2011), pp. 1254–1272 (cit. on p. 50).

[199] Nikola T Markov, Julien Vezoli, Pascal Chameau, et al. "Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex". In: *Journal of Comparative Neurology* 522.1 (2014), pp. 225–259 (cit. on pp. 50, 52, 171, 172).

[200] David Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. 1982 (cit. on pp. 19, 22).

[201] Kevan AC Martin. "A brief history of the "feature detector"". In: *Cerebral cortex* 4.1 (1994), pp. 1–7 (cit. on pp. 56, 57, 171, 172).

[202] Richard H Masland. "The neuronal organization of the retina". In: *Neuron* 76.2 (2012), pp. 266–280 (cit. on p. 27).

[203] Abraham Harold Maslow. "The psychology of science: A reconnaissance". In: (1966) (cit. on p. 52).

[204] Ron McClamrock. "Marr's three levels: A re-evaluation". In: *Minds and Machines* 1 (1991), pp. 185–196 (cit. on p. 22).

[205] Lotfi Merabet, Alex Desautels, Karine Minville, et al. "Motion integration in a thalamic visual nucleus". In: *Nature* 396.6708 (1998), pp. 265–268 (cit. on p. 133).

[206] David Meunier, Renaud Lambiotte, and Edward T Bullmore. "Modular and hierarchically modular organization of brain networks". In: *Frontiers in neuroscience* 4 (2010), p. 200 (cit. on pp. 120, 128).

[207] Travis Meyer and Carl R Olson. "Statistical learning of visual transitions in monkey inferotemporal cortex". In: *Proceedings of the National Academy of Sciences* 108.48 (2011), pp. 19401–19406 (cit. on p. 52).

[208] Ethan M Meyers. "The neural decoding toolbox". In: *Frontiers in neuroinformatics* 7 (2013), p. 8 (cit. on p. 97).

[209] René Michel, Laura Dugué, and Niko A Busch. "Distinct contributions of alpha and theta rhythms to perceptual and attentional sampling". In: *European Journal of Neuroscience* 55.11-12 (2022), pp. 3025–3039 (cit. on p. 165).

[210] David Millett. "Hans Berger: From psychic energy to the EEG". In: *Perspectives in biology and medicine* 44.4 (2001), pp. 522–542 (cit. on p. 96).

[211] Beren Millidge, Anil Seth, and Christopher L Buckley. "Predictive coding: a theoretical and experimental review". In: *arXiv preprint arXiv:2107.12979* (2021) (cit. on pp. 43, 49, 133).

[212] Beren Millidge, Alexander Tschantz, and Christopher L Buckley. "Predictive coding approximates backprop along arbitrary computation graphs". In: *Neural Computation* 34.6 (2022), pp. 1329–1368 (cit. on pp. 47, 166).

[213] Beren Millidge, Alexander Tschantz, Anil Seth, et al. "Relaxing the constraints on predictive coding models". In: *arXiv preprint arXiv:2010.01047* (2020) (cit. on pp. 44, 49, 166).

[214] Cyril Monier, Frederic Chavane, Pierre Baudot, et al. "Orientation and direction selectivity of synaptic inputs in visual cortical neurons: a diversity of combinations produces spike tuning". In: *Neuron* 37.4 (2003), pp. 663–680 (cit. on p. 55).

[215] Rosalyn J Moran, Pablo Campo, Mkael Symmonds, et al. "Free energy, precision and learning: the role of cholinergic neuromodulation". In: *Journal of Neuroscience* 33.19 (2013), pp. 8227–8236 (cit. on pp. 160, 162).

[216] Vernon B Mountcastle. "The columnar organization of the neocortex." In: *Brain: a journal of neurology* 120.4 (1997), pp. 701–722 (cit. on p. 28).

[217] Lyle Muller, Frédéric Chavane, John Reynolds, et al. "Cortical travelling waves: mechanisms and computational principles". In: *Nature Reviews Neuroscience* 19.5 (2018), pp. 255–268 (cit. on pp. 165, 168, 171).

[218] Kathryn M Murphy, David G Jones, Suzanne B Fenstemaker, et al. "Spacing of cytochrome oxidase blobs in visual cortex of normal and strabismic monkeys." In: *Cerebral cortex (New York, NY: 1991)* 8.3 (1998), pp. 237–244 (cit. on p. 29).

[219] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012 (cit. on pp. 41, 165).

[220] PC Murphy and AM Sillito. "Corticofugal feedback influences the generation of length tuning in the visual pathway". In: *Nature* 329.6141 (1987), pp. 727–729 (cit. on p. 93).

[221] Scott O Murray, Daniel Kersten, Bruno A Olshausen, et al. "Shape perception reduces activity in human primary visual cortex". In: *Proceedings of the National Academy of Sciences* 99.23 (2002), pp. 15164–15169 (cit. on p. 52).

[222] Alexander Naka and Hillel Adesnik. "Inhibitory circuits in cortical layer 5". In: *Frontiers in neural circuits* 10 (2016), p. 35 (cit. on p. 51).

[223] Ken-Ichi Naka and William AH Rushton. "S-potentials from luminosity units in the retina of fish (Cyprinidae)". In: *The Journal of physiology* 185.3 (1966), pp. 587–599 (cit. on p. 31).

[224] Erwin Neher and Bert Sakmann. "The patch clamp technique". In: *Scientific American* 266.3 (1992), pp. 44–51 (cit. on p. 94).

[225] Eric Newman and Andreas Reichenbach. "The Müller cell: a functional element of the retina". In: *Trends in neurosciences* 19.8 (1996), pp. 307–312 (cit. on p. 26).

[226] Amy M Ni, Douglas A Ruff, Joshua J Alberts, et al. "Learning and attention reveal a general relationship between population activity and behavior". In: *Science* 359.6374 (2018), pp. 463–465 (cit. on p. 117).

[227] Danko Nikolić, Pascal Fries, and Wolf Singer. "Gamma oscillations: precise temporal coordination without a metronome". In: *Trends in cognitive sciences* 17.2 (2013), pp. 54–55 (cit. on p. 51).

[228] Simon Nougaret, Laura Lopez-Galdo, Emile Caytan, et al. "Distinct sources and behavioral correlates of macaque motor cortical low and high beta". In: *bioRxiv* (2023), pp. 2023–03 (cit. on p. 97).

[229] John O'Keefe and Jonathan Dostrovsky. "The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat." In: *Brain research* (1971) (cit. on pp. 56, 163).

[230] Bruno A Olshausen and David J Field. "Emergence of simple-cell receptive field properties by learning a sparse code for natural images". In: *Nature* 381.6583 (1996), pp. 607–609 (cit. on pp. 33, 55, 57, 59).

[231] Bruno A Olshausen and David J Field. "Natural image statistics and efficient coding". In: *Network: computation in neural systems* 7.2 (1996), p. 333 (cit. on p. 31).

[232] Bruno A Olshausen and David J Field. "Sparse coding with an overcomplete basis set: A strategy employed by V1?" In: *Vision research* 37.23 (1997), pp. 3311–3325 (cit. on pp. 33, 57).

[233] Bence P Ölveczky, Stephen A Baccus, and Markus Meister. "Segregation of object and background motion in the retina". In: *Nature* 423.6938 (2003), pp. 401–408 (cit. on p. 27).

[234] Guy A Orban. *Neuronal operations in the visual cortex*. Vol. 11. Springer Science & Business Media, 2012 (cit. on p. 28).

[235] Gergő Orbán, Pietro Berkes, József Fiser, et al. "Neural variability and sampling-based probabilistic representations in the visual cortex". In: *Neuron* 92.2 (2016), pp. 530–543 (cit. on p. 54).

[236] Marius Pachitariu, Nicholas Steinmetz, Shabnam Kadir, et al. "Kilosort: real-time spike-sorting for extracellular electrophysiology with hundreds of channels". In: *BioRxiv* (2016), p. 061481 (cit. on p. 95).

[237] Alexandr Pak, Samuel T Kissinger, and Alexander A Chubykin. "Impaired adaptation and laminar processing of the oddball paradigm in the primary visual cortex of Fmr1 KO mouse". In: *Frontiers in Cellular Neuroscience* 15 (2021), p. 668230 (cit. on p. 52).

[238] Benjamin A Palmer, Gavin J Taylor, Vlad Brumfeld, et al. "The image-forming mirror in the eye of the scallop". In: *Science* 358.6367 (2017), pp. 1172–1175 (cit. on p. 24).

[239] Adam Paszke, Sam Gross, Soumith Chintala, et al. "Automatic differentiation in pytorch". In: (2017) (cit. on pp. 85, 166).

[240] Bertram Payne and Alan Peters. *The cat primary visual cortex*. Academic Press, 2001 (cit. on pp. 29, 31).

[241] David Peebles and Richard P Cooper. *Thirty years after Marr's vision: Levels of analysis in cognitive science*. 2015 (cit. on p. 22).

[242] Jacob Penner, Elizabeth A Osuch, Betsy Schaefer, et al. "Higher order thalamic nuclei resting network connectivity in early schizophrenia and major depressive disorder". In: *Psychiatry Research: Neuroimaging* 272 (2018), pp. 7–16 (cit. on p. 163).

[243] David I Perrett, Edmond T Rolls, and Woody Caan. "Visual neurones responsive to faces in the monkey temporal cortex". In: *Experimental brain research* 47 (1982), pp. 329–342 (cit. on p. 56).

[244] Laurent U Perrinet, Rick A Adams, and Karl J Friston. "Active inference, eye movements and oculomotor delays". In: *Biological cybernetics* 108 (2014), pp. 777–801 (cit. on p. 45).

[245] Laurent U Perrinet and James A Bednar. "Edge co-occurrences can account for rapid categorization of natural versus animal images". In: *Scientific reports* 5.1 (2015), p. 11400 (cit. on pp. 33, 57).

[246] Aaron A Phillips, Franco HN Chan, Mei Mu Zi Zheng, et al. "Neurovascular coupling in humans: physiology, methodological advances and clinical implications". In: *Journal of Cerebral Blood Flow & Metabolism* 36.4 (2016), pp. 647–664 (cit. on p. 96).

[247] Gregory C Phillips and Hugh R Wilson. "Orientation bandwidths of spatial mechanisms measured by masking". In: *JOSA A* 1.2 (1984), pp. 226–232 (cit. on p. 36).

[248] Jasper Poort, Florian Raudies, Aurel Wannig, et al. "The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex". In: *Neuron* 75.1 (2012), pp. 143–156 (cit. on pp. 92, 93).

[249] Nicholas J Priebe. "Mechanisms of orientation selectivity in the primary visual cortex". In: *Annual review of vision science* 2 (2016), pp. 85–107 (cit. on p. 33).

[250] Nicholas J Priebe and David Ferster. "Mechanisms of neuronal computation in mammalian visual cortex". In: *Neuron* 75.2 (2012), pp. 194–208 (cit. on pp. 90, 91).

[251] Nicholas J Priebe, Stephen G Lisberger, and J Anthony Movshon. "Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex". In: *Journal of Neuroscience* 26.11 (2006), pp. 2941–2950 (cit. on p. 117).

[252] Gopathy Purushothaman, Roan Marion, Keji Li, et al. "Gating and control of primary visual cortex by pulvinar". In: *Nature neuroscience* 15.6 (2012), pp. 905–912 (cit. on pp. 53, 133, 135, 160).

[253] Mostafa Rahimi Azghadi, Ying-Chen Chen, Jason K Eshraghian, et al. "Complementary metal-oxide semiconductor and memristive hardware for neuromorphic computing". In: *Advanced Intelligent Systems* 2.5 (2020), p. 1900189 (cit. on p. 167).

[254] Rajesh PN Rao and Dana H Ballard. "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects". In: *Nature neuroscience* 2.1 (1999), pp. 79–87 (cit. on pp. 44, 49, 169).

[255] Cesar R Ravello, Laurent U Perrinet, Maria-Jose Escobar, et al. "Speed-selectivity in retinal ganglion cells is sharpened by broad spatial frequency, naturalistic stimuli". In: *Scientific reports* 9.1 (2019), pp. 1–16 (cit. on p. 35).

[256] Dario L Ringach, Michael J Hawken, and Robert Shapley. "Dynamics of orientation tuning in macaque primary visual cortex". In: *Nature* 387.6630 (1997), pp. 281–284 (cit. on p. 55).

[257] Dario L Ringach, Robert M Shapley, and Michael J Hawken. "Orientation selectivity in macaque V1: diversity and laminar dependence". In: *Journal of neuroscience* 22.13 (2002), pp. 5639–5651 (cit. on pp. 36, 55, 57, 112).

[258] David Lee Robinson and Steven E Petersen. "The pulvinar and visual salience". In: *Trends in neurosciences* 15.4 (1992), pp. 127–132 (cit. on pp. 53, 133, 160).

[259] Klaus Rohrschneider. "Determination of the location of the fovea on the fundus". In: *Investigative ophthalmology & visual science* 45.9 (2004), pp. 3257–3258 (cit. on p. 26).

[260] Cyrille Rossant, Shabnam N Kadir, Dan FM Goodman, et al. "Spike sorting for large, dense electrode arrays". In: *Nature neuroscience* 19.4 (2016), pp. 634–641 (cit. on p. 95).

[261] Christopher J Rozell, Don H Johnson, Richard G Baraniuk, et al. "Sparse coding via thresholding and local competition in neural circuits". In: *Neural computation* 20.10 (2008), pp. 2526–2563 (cit. on p. 86).

[262] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. "Learning representations by back-propagating errors". In: *nature* 323.6088 (1986), pp. 533–536 (cit. on p. 61).

[263] Nicole C Rust and J Anthony Movshon. "In praise of artifice". In: *Nature neuroscience* 8.12 (2005), pp. 1647–1650 (cit. on pp. 19, 33–35).

[264] João Sacramento, Rui Ponte Costa, Yoshua Bengio, et al. "Dendritic cortical microcircuits approximate the backpropagation algorithm". In: *Advances in neural information processing systems* 31 (2018) (cit. on p. 51).

[265] Bert Sakmann and Erwin Neher. "Patch clamp techniques for studying ionic channels in excitable membranes". In: *Annual review of physiology* 46.1 (1984), pp. 455–472 (cit. on p. 94).

[266] Remi Sanchez, Karen Davranche, Thibault Gajdos, et al. "Action monitoring boosts perceptual confidence". In: *bioRxiv* (2023), pp. 2023–08 (cit. on p. 36).

[267] Jack W Scannell, Colin Blakemore, and Malcolm P Young. "Analysis of connectivity in the cat cerebral cortex". In: *Journal of Neuroscience* 15.2 (1995), pp. 1463–1483 (cit. on p. 30).

[268] Markus K Schaefer, Julio C Hechavarria, and Manfred Kössl. "Quantification of mid and late evoked sinks in laminar current source density profiles of columns in the primary auditory cortex". In: *Frontiers in neural circuits* 9 (2015), p. 52 (cit. on p. 96).

[269] Kerstin E Schmidt and Fred Wolf. "Punctuated evolution of visual cortical circuits? Evidence from the large rodent Dasyprocta leporina, and the tiny primate Microcebus murinus". In: *Current Opinion in Neurobiology* 71 (2021), pp. 110–118 (cit. on p. 31).

[270] Joscha T Schmiedt, Alexander Maier, Pascal Fries, et al. "Beta oscillation dynamics in extrastriate cortex after removal of primary visual cortex". In: *Journal of Neuroscience* 34.35 (2014), pp. 11857–11864 (cit. on p. 51).

[271] Matthew Schmolesky. "The primary visual cortex". In: (2011) (cit. on p. 29).

[272] Steffen Schneider, Jin Hwa Lee, and Mackenzie Weygandt Mathis. "Learnable latent embeddings for joint behavioural and neural analysis". In: *Nature* (2023), pp. 1–9 (cit. on p. 114).

[273] Benjamin Scholl, Daniel E Wilson, and David Fitzpatrick. "Local order within global disorder: synaptic architecture of visual space". In: *Neuron* 96.5 (2017), pp. 1127–1138 (cit. on p. 55).

[274] Wolfram Schultz, Peter Dayan, and P Read Montague. "A neural substrate of prediction and reward". In: *Science* 275.5306 (1997), pp. 1593–1599 (cit. on p. 164).

[275] Mehdi Senoussi, James C Moreland, Niko A Busch, et al. "Attention explores space periodically at the theta frequency". In: *Journal of Vision* 19.5 (2019), pp. 22–22 (cit. on p. 165).

[276] Thomas Serre and Tomaso Poggio. "A neuromorphic approach to computer vision". In: *Communications of the ACM* 53.10 (2010), pp. 54–61 (cit. on p. 88).

[277] Claude Elwood Shannon. "A mathematical theory of communication". In: *The Bell system technical journal* 27.3 (1948), pp. 379–423 (cit. on p. 42).

[278] Gordon M Shepherd. "The microcircuit concept applied to cortical evolution: from three-layer to six-layer cortex". In: *Frontiers in neuroanatomy* 5 (2011), p. 30 (cit. on p. 171).

[279] S Murray Sherman. "The thalamus is more than just a relay". In: *Current opinion in neurobiology* 17.4 (2007), pp. 417–422 (cit. on p. 27).

[280] S Murray Sherman and RW Guillery. "On the actions that one nerve cell can have on another: distinguishing "drivers" from "modulators"". In: *Proceedings of the National Academy of Sciences* 95.12 (1998), pp. 7121–7126 (cit. on p. 134).

[281] Charles Scott Sherrington. "Observations on the scratch-reflex in the spinal dog". In: *The Journal of physiology* 34.1-2 (1906), p. 1 (cit. on pp. 28, 56, 163).

[282] S Shipp. "The functional logic of cortico–pulvinar connections". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358.1438 (2003), pp. 1605–1624 (cit. on p. 133).

[283] Stewart Shipp. "Neural elements for predictive coding". In: *Frontiers in psychology* 7 (2016), p. 1792 (cit. on pp. 49, 52, 169).

[284] Stewart Shipp, Rick A Adams, and Karl J Friston. "Reflections on agranular architecture: predictive coding in the motor cortex". In: *Trends in neurosciences* 36.12 (2013), pp. 706–716 (cit. on p. 28).

[285] R Angus Silver. "Neuronal arithmetic". In: *Nature Reviews Neuroscience* 11.7 (2010), pp. 474–489 (cit. on p. 134).

[286] Eero P Simoncelli and Bruno A Olshausen. "Natural image statistics and neural representation". In: *Annual review of neuroscience* 24.1 (2001), pp. 1193–1216 (cit. on p. 31).

[287] Claudio Simoncini, Laurent U Perrinet, Anna Montagnini, et al. "More is not always better: adaptive gain control explains dissociation between perception and action". In: *Nature neuroscience* 15.11 (2012), pp. 1596–1603 (cit. on p. 35).

[288] Bernt C Skottun. "A model for end-stopping in the visual cortex". In: *Vision research* 38.13 (1998), pp. 2023–2035 (cit. on p. 49).

[289] Bernt C Skottun, Russell L De Valois, David H Grosof, et al. "Classifying simple and complex cells on the basis of response modulation". In: *Vision research* 31.7-8 (1991), pp. 1078–1086 (cit. on p. 112).

[290] Stephen Smith, Til Ole Bergmann, Birte Forstmann, et al. "Imaging Neuroscience opening editorial". In: *Imaging Neuroscience* (2023) (cit. on p. 224).

[291] Robert J Snowden. "Orientation bandwidth: The effect of spatial and temporal frequency". In: *Vision research* 32.10 (1992), pp. 1965–1974 (cit. on p. 36).

[292] Bruno Oliveira Ferreira de Souza, Nelson Cortes, and Christian Casanova. "Pulvinar modulates contrast responses in the visual cortex as a function of cortical hierarchy". In: *Cerebral Cortex* 30.3 (2020), pp. 1068–1086 (cit. on pp. 20, 53).

[293] Michael W Spratling. "A hierarchical predictive coding model of object recognition in natural images". In: *Cognitive computation* 9.2 (2017), pp. 151–167 (cit. on p. 44).

[294] Mark Sprevak. "Predictive coding II: The computational level". In: (2021) (cit. on p. 173).

[295] Nicholas A Steinmetz, Cagatay Aydin, Anna Lebedeva, et al. "Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings". In: *Science* 372.6539 (2021), eabf4588 (cit. on p. 95).

[296] Nicholas A Steinmetz, Christof Koch, Kenneth D Harris, et al. "Challenges and opportunities for large-scale electrophysiology with Neuropixels probes". In: *Current opinion in neurobiology* 50 (2018), pp. 92–100 (cit. on p. 95).

[297] Peter Sterling. "Microcircuitry of the cat retina". In: *Annual review of neuroscience* 6.1 (1983), pp. 149–185 (cit. on p. 30).

[298] Peter Sterling and Simon Laughlin. *Principles of neural design*. MIT press, 2015 (cit. on p. 130).

[299] Philipp Sterzer, Rick A Adams, Paul Fletcher, et al. "The predictive coding account of psychosis". In: *Biological psychiatry* 84.9 (2018), pp. 634–643 (cit. on p. 163).

[300] Peter S Stevens. "Patterns in nature". In: (1974) (cit. on p. 31).

[301] JONATHAN Stone and BOGDAN Dreher. "Projection of X-and Y-cells of the cat's lateral geniculate nucleus to areas 17 and 18 of visual cortex." In: *Journal of Neurophysiology* 36.3 (1973), pp. 551–567 (cit. on p. 31).

[302] Christoph Stosiek, Olga Garashuk, Knut Holthoff, et al. "In vivo two-photon calcium imaging of neuronal networks". In: *Proceedings of the National Academy of Sciences* 100.12 (2003), pp. 7319–7324 (cit. on p. 96).

[303] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, et al. "High-dimensional geometry of population responses in visual cortex". In: *Nature* 571.7765 (2019), pp. 361–365 (cit. on p. 96).

[304] Christopher Summerfield, Emily H Trittschuh, Jim M Monti, et al. "Neural repetition suppression reflects fulfilled perceptual expectations". In: *Nature neuroscience* 11.9 (2008), pp. 1004–1006 (cit. on p. 52).

[305] Christopher Summerfield, Valentin Wyart, Vanessa Mareike Johnen, et al. "Human scalp electroencephalography reveals that repetition suppression varies with expectation". In: *Frontiers in Human Neuroscience* 5 (2011), p. 67 (cit. on p. 52).

[306] Hans Super and Pieter R Roelfsema. "Chronic multiunit recordings in behaving animals: advantages and limitations". In: *Progress in brain research* 147 (2005), pp. 263–282 (cit. on p. 118).

[307] Mototaka Suzuki and Matthew E Larkum. "General anesthesia decouples cortical pyramidal neurons". In: *Cell* 180.4 (2020), pp. 666–676 (cit. on pp. 167, 171).

[308] Nicholas V Swindale. "Orientation tuning curves: empirical description and estimation of parameters". In: *Biological cybernetics* 78.1 (1998), pp. 45–56 (cit. on p. 34).

[309] Richard Szeliski. "Bayesian modeling of uncertainty in low-level vision". In: *International Journal of Computer Vision* 5.3 (1990), pp. 271–301 (cit. on p. 36).

[310] Toru Takahata. "What does cytochrome oxidase histochemistry represent in the visual cortex?" In: *Frontiers in neuroanatomy* 10 (2016), p. 79 (cit. on p. 29).

[311] Andrew YY Tan, Brandon D Brown, Benjamin Scholl, et al. "Orientation selectivity of synaptic input to neurons in mouse and cat primary visual cortex". In: *Journal of Neuroscience* 31.34 (2011), pp. 12339–12350 (cit. on pp. 90, 91).

[312] Hanne H Thoen, Martin J How, Tsyr-Huei Chiou, et al. "A different form of color vision in mantis shrimp". In: *Science* 343.6169 (2014), pp. 411–413 (cit. on p. 24).

[313] Alex M Thomson and A Peter Bannister. "Interlaminar connections in the neocortex". In: *Cerebral cortex* 13.1 (2003), pp. 5–14 (cit. on p. 52).

[314] Alex M Thomson, David C West, Yun Wang, et al. "Synaptic connections and small circuits involving excitatory and inhibitory neurons in layers 2–5 of adult rat and cat neocortex: triple intracellular recordings and biocytin labelling in vitro". In: *Cerebral cortex* 12.9 (2002), pp. 936–953 (cit. on p. 50).

[315] Simon Thorpe, Denis Fize, and Catherine Marlot. "Speed of processing in the human visual system". In: *nature* 381.6582 (1996), pp. 520–522 (cit. on p. 31).

[316] Ana Todorovic, Freek van Ede, Eric Maris, et al. "Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study". In: *Journal of Neuroscience* 31.25 (2011), pp. 9118–9123 (cit. on p. 52).

[317] David J Tolhurst, Yoav Tadmor, and Tang Chao. "Amplitude spectra of natural images". In: *Ophthalmic and Physiological Optics* 12.2 (1992), pp. 229–232 (cit. on p. 33).

[318] DJ Tolhurst and ID Thompson. "On the variety of spatial frequency selectivities shown by neurons in area 17 of the cat". In: *Proceedings of the Royal Society of London. Series B. Biological Sciences* 213.1191 (1981), pp. 183–199 (cit. on p. 29).

[319] V Javier Traver and Alexandre Bernardino. "A review of log-polar imaging for visual perception in robotics". In: *Robotics and Autonomous Systems* 58.4 (2010), pp. 378–398 (cit. on p. 30).

[320] Todd W Troyer, Anton E Krukowski, Nicholas J Priebe, et al. "Contrast-invariant orientation tuning in cat visual cortex: thalamocortical input tuning and correlation-based intracortical connectivity". In: *Journal of Neuroscience* 18.15 (1998), pp. 5908–5927 (cit. on p. 169).

[321] Jonathan Vacher. "Synthese de textures dynamiques pour l'etude de la vision en psychophysique et electrophysiologie". PhD thesis. École Normale Supérieure, 2017 (cit. on p. 35).

[322] Jonathan Vacher, Andrew Isaac Meso, Laurent U Perrinet, et al. "Biologically inspired dynamic textures for probing motion perception". In: *Advances in neural information processing systems* 28 (2015) (cit. on p. 34).

[323] Jeroen JA Van Boxtel and Hongjing Lu. *A predictive coding perspective on autism spectrum disorders*. 2013 (cit. on p. 53).

[324] Sander Van de Cruys, Kris Evers, Ruth Van der Hallen, et al. "Precise minds in uncertain worlds: predictive coding in autism." In: *Psychological review* 121.4 (2014), p. 649 (cit. on pp. 53, 163).

[325] David C Van Essen, Charles H Anderson, and Daniel J Felleman. "Information processing in the primate visual system: an integrated systems perspective". In: *Science* 255.5043 (1992), pp. 419–423 (cit. on p. 87).

[326] Stephen D Van Hooser, Arani Roy, Heather J Rhodes, et al. "Transformation of receptive field properties from lateral geniculate nucleus to superficial V1 in the tree shrew". In: *Journal of Neuroscience* 33.28 (2013), pp. 11494–11505 (cit. on p. 90).

[327] E Vaucher, D Linville, and E Hamel. "Cholinergic basal forebrain projections to nitric oxide synthase-containing neurons in the rat cerebral cortex". In: *Neuroscience* 79.3 (1997), pp. 827–836 (cit. on p. 164).

[328] Mateo Vélez-Fort, Charly V Rousseau, Christian J Niedworok, et al. "The stimulus selectivity and connectivity of layer six principal cells reveals cortical microcircuits underlying visual processing". In: *Neuron* 83.6 (2014), pp. 1431–1443 (cit. on p. 167).

[329] TR Vidyasagar, X Pei, and M Volgushev. "Multiple mechanisms underlying the orientation selectivity of visual cortical neurones". In: *Trends in neurosciences* 19.7 (1996), pp. 272–277 (cit. on p. 91).

[330] Trichur R Vidyasagar and Ulf T Eysel. "Origins of feature selectivities and maps in the mammalian primary visual cortex". In: *Trends in neurosciences* 38.8 (2015), pp. 475–485 (cit. on pp. 90, 91, 169).

[331] Martin Y Villeneuve and Christian Casanova. "On the use of isoflurane versus halothane in the study of visual response properties of single cells in the primary visual cortex". In: *Journal of neuroscience methods* 129.1 (2003), pp. 19–31 (cit. on p. 167).

[332] MY Villeneuve, B Thompson, RF Hess, et al. "Pattern-motion selective responses in MT, MST and the pulvinar of humans". In: *European Journal of Neuroscience* 36.6 (2012), pp. 2849–2858 (cit. on p. 133).

[333] Erich Von Holst and Horst Mittelstaedt. "Das reafferenzprinzip: wechselwirkungen zwischen zentralnervensystem und peripherie". In: *Naturwissenschaften* 37.20 (1950), pp. 464–476 (cit. on p. 88).

[334] Catherine Wacongne, Jean-Pierre Changeux, and Stanislas Dehaene. "A neuronal model of predictive coding accounting for the mismatch negativity". In: *Journal of Neuroscience* 32.11 (2012), pp. 3665–3678 (cit. on p. 52).

[335] Brian A Wandell, Serge O Dumoulin, and Alyssa A Brewer. "Visual field maps in human cortex". In: *Neuron* 56.2 (2007), pp. 366–383 (cit. on p. 30).

[336] Duncan J Watts and Steven H Strogatz. "Collective dynamics of 'small-world' networks". In: *nature* 393.6684 (1998), pp. 440–442 (cit. on p. 123).

[337] Haiguang Wen, Kuan Han, Junxing Shi, et al. "Deep predictive coding network for object recognition". In: *International conference on machine learning*. PMLR. 2018, pp. 5266–5275 (cit. on p. 44).

[338]   John Simon Werner and Leo M Chalupa. *The visual neurosciences.* Mit Press, 2004 (cit. on p. 133).

[339]   Brendt Wohlberg. "SPORCO: A Python package for standard and convolutional sparse representations". In: *Proceedings of the 15th Python in Science Conference, Austin, TX, USA.* 2017, pp. 1–8 (cit. on p. 85).

[340]   Keith J Worsley and Karl J Friston. "Analysis of fMRI time-series revisited—again". In: *Neuroimage* 2.3 (1995), pp. 173–181 (cit. on p. 96).

[341]   Christian Wozny and Stephen R Williams. "Specificity of synaptic connectivity between layer 1 inhibitory interneurons and layer 2/3 pyramidal neurons in the rat neocortex". In: *Cerebral cortex* 21.8 (2011), pp. 1818–1826 (cit. on p. 93).

[342]   Xiangmin Xu, Jennifer Ichida, Yuri Shostak, et al. "Are primate lateral geniculate nucleus (LGN) cells really sensitive to orientation or direction?" In: *Visual neuroscience* 19.1 (2002), pp. 97–108 (cit. on p. 90).

[343]   Maeri Yamamoto, Itaru Kushima, Ryohei Suzuki, et al. "Aberrant functional connectivity between the thalamus and visual cortex is related to attentional impairment in schizophrenia". In: *Psychiatry Research: Neuroimaging* 278 (2018), pp. 35–41 (cit. on p. 163).

[344]   Takashi Yoshida and Kenichi Ohki. "Natural images are reliably represented by sparse and variable populations of neurons in visual cortex". In: *Nature communications* 11.1 (2020), pp. 1–19 (cit. on p. 33).

[345]   John Zachary Young. "The optic lobes of Octopus vulgaris". In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 245.718 (1962), pp. 19–58 (cit. on p. 24).

APPENDICES

A. Appendix A: Additional Equations

A.1. Equation 2.9

The full differentiation of F over Φ serves no introductory purpose, and is thus left out of Chapter 2 - equation 2.9. It is written here, as:

$$\frac{\delta F}{\delta \Phi} = \frac{1}{2}\left(\frac{\delta}{\delta \Phi}\left(-\frac{(u-g(\Phi))^2}{\Sigma_u}\right) + \frac{\delta}{\delta \Phi}\left(-\frac{(\Phi-v_p)^2}{\Sigma_p}\right) + \frac{\delta}{\delta \Phi}\left(-\ln\Sigma_u\right) + \frac{\delta}{\delta \Phi}\left(-\ln\Sigma_p\right) + \frac{\delta}{\delta \Phi}C\right)$$

$$= \frac{1}{2}\left(\left(-\frac{1}{\Sigma_u}\frac{\delta}{\delta \Phi}(u-g(\Phi))^2\right) + \left(-\frac{1}{\Sigma_p}\frac{\delta}{\delta \Phi}(\Phi-v_p)^2\right)\right)$$

Applying the power rule $(f(x)^n)' = n f(x)^{n-1} f'(x)$:

$$= \frac{1}{2}\left(\left(-\frac{1}{\Sigma_u}2(u-g(\Phi))\frac{\delta}{\delta \Phi}(u-g(\Phi))\right) + \left(-\frac{1}{\Sigma_p}2(\Phi-v_p)\frac{\delta}{\delta \Phi}(\Phi-v_p)\right)\right)$$

And then splitting the linear differentiation:

$$= \frac{1}{2}\left(\left(-\frac{1}{\Sigma_u}2(u-g(\Phi))\left(\frac{\delta}{\delta \Phi}u - \frac{\delta}{\delta \Phi}g(\Phi)\right)\right) + \left(-\frac{1}{\Sigma_p}2(\Phi-v_p)\left(\frac{\delta}{\delta \Phi}\Phi - \frac{\delta}{\delta \Phi}v_p\right)\right)\right)$$

$$= \frac{1}{2}\left(\left(-\frac{1}{\Sigma_u}2(u-g(\Phi))\left(0-g'(\Phi)\right)\right) + \left(-\frac{1}{\Sigma_p}2(\Phi-v_p)(1+0)\right)\right)$$

$$= \frac{1}{2}\left(\left(-\frac{1}{\Sigma_u}2(u-g(\Phi))(-g'(\Phi))\right) + \left(-\frac{1}{\Sigma_p}2(\Phi-v_p)\right)\right)$$

$$= \left(\frac{1}{\Sigma_u}(u-g(\Phi))(g'(\Phi))\right) + \left(-\frac{1}{\Sigma_p}(\Phi-v_p)\right)$$

Simplifying and changing the order of the right-hand side term, we get:

$$F = \frac{(u-g(\Phi))}{\Sigma_u}g'(\Phi) + \frac{(v_p-\Phi)}{\Sigma_p}$$

yielding equation 2.9.

A.2. Equation 2.25

For differentiating the terms $v_p, \Sigma_u \Sigma_p$ over F for Equation 2.25, we start from the definition of F as:

$$F = \frac{1}{2}\left(-\ln\Sigma_p - \frac{(\Phi - v_p)^2}{\Sigma_p} - \ln\Sigma_u - \frac{(u - g(\Phi))^2}{\Sigma_u}\right) + C$$

We will derive first for v_p:

$$\frac{\delta F}{\delta v_p} = \frac{1}{2}\left(\frac{\delta}{\delta v_p}\left(-\frac{(\Phi - v_p)^2}{\Sigma_p}\right) + \frac{\delta}{\delta v_p}\left(-\frac{(u - g(\Phi))^2}{\Sigma_u}\right) + \frac{\delta}{\delta v_p}(-\ln\Sigma_u) + \frac{\delta}{\delta v_p}(-\ln\Sigma_p) +\right)\frac{\delta}{\delta v_p}C$$

$$= \frac{1}{2}\left(\frac{\delta}{\delta v_p}\left(-\frac{(\Phi - v_p)^2}{\Sigma_p}\right)\right)$$

$$= \frac{1}{2}\left(\frac{1}{-\Sigma_p}\frac{\delta}{\delta v_p}(\Phi - v_p)^2\right)$$

Using the power rule to eliminate both halved and squared terms:

$$= \frac{1}{-\Sigma_p}(\Phi - v_p)\frac{\delta}{\delta v_p}(\Phi - v_p)$$

$$= \frac{1}{-\Sigma_p}(\Phi - v_p)\left(\frac{\delta}{\delta v_p}\Phi - \frac{\delta}{\delta v_p}v_p\right)$$

$$= \frac{1}{-\Sigma_p}(\Phi - v_p)(0 - 1)$$

$$= \frac{\Phi - v_p}{\Sigma_p}$$

Now once again, for Σ_p

$$F = \frac{1}{2}\left(-\ln\Sigma_p - \frac{(\Phi - v_p)^2}{\Sigma_p} - \ln\Sigma_u - \frac{(u - g(\Phi))^2}{\Sigma_u}\right) + C$$

$$\frac{\delta F}{\delta \Sigma_p} = \frac{1}{2}\left(\frac{\delta}{\delta \Sigma_p}\left(-\frac{(\Phi - v_p)^2}{\Sigma_p}\right) + \frac{\delta}{\delta \Sigma_p}\left(-\frac{(u - g(\Phi))^2}{\Sigma_u}\right) + \frac{\delta}{\delta \Sigma_p}(-\ln\Sigma_u) + \frac{\delta}{\delta \Sigma_p}(-\ln\Sigma_p)\right) + \frac{\delta}{\delta \Sigma_p}C$$

$$= \frac{1}{2}\left(-\frac{\delta}{\delta \Sigma_p}\ln\Sigma_p + \left(-(\Phi - v_p)^2\frac{\delta}{\delta \Sigma_p}\frac{1}{\Sigma_p}\right)\right)$$

Applying $\left(\frac{1}{f(x)}\right)' = -\frac{f'(x)}{f(x)^2}$, this becomes

$$= \frac{1}{2}\left(-\frac{1}{\Sigma_p} + \left((\Phi - v_p)^2 \frac{\frac{\delta}{\delta\Sigma_p}\Sigma_p}{\Sigma_p^2}\right)\right)$$

$$= \frac{1}{2}\left((\Phi - v_p)^2 \frac{\frac{\delta}{\delta\Sigma_p}\Sigma_p}{\Sigma_p^2} - \frac{1}{\Sigma_p}\right)$$

$$= \frac{1}{2}\left(\frac{1(\Phi - v_p)^2}{\Sigma_p^2} - \frac{1}{\Sigma_p}\right)$$

$$= \frac{1}{2}\left(\frac{(\Phi - v_p)^2}{\Sigma_p^2} - \frac{1}{\Sigma_p}\right)$$

The same is done for Σ_u:

$$F = \frac{1}{2}\left(-\ln\Sigma_p - \frac{(\Phi - v_p)^2}{\Sigma_p} - \ln\Sigma_u - \frac{(u - g(\Phi))^2}{\Sigma_u}\right) + C$$

$$\frac{\delta F}{\delta\Sigma_u} = \frac{1}{2}\left(\frac{\delta}{\delta\Sigma_u}\left(-\frac{(\Phi - v_p)^2}{\Sigma_p}\right) + \frac{\delta}{\delta\Sigma_u}\left(-\frac{(u - g(\Phi))^2}{\Sigma_u}\right) + \frac{\delta}{\delta\Sigma_u}\left(-\ln\Sigma_u\right) + \frac{\delta}{\delta\Sigma_u}\left(-\ln\Sigma_p\right)\right) + \frac{\delta}{\delta\Sigma_u}C$$

$$= \frac{1}{2}\left(\frac{\delta}{\delta\Sigma_u}\left(-\frac{(u - g(\Phi))^2}{\Sigma_u}\right) - \frac{\delta}{\delta\Sigma_u}\ln\Sigma_u\right)$$

$$= \frac{1}{2}\left(-(u - g(\Phi))^2 \frac{\delta}{\delta\Sigma_u}\frac{1}{\Sigma_u} - \frac{\delta}{\delta\Sigma_u}\ln\Sigma_u\right)$$

$$= \frac{1}{2}\left(-(u - g(\Phi))^2 \frac{\delta}{\delta\Sigma_u}\frac{1}{\Sigma_u} - \frac{1}{\Sigma_u}\right)$$

Once more, using $\left(\frac{1}{f(x)}\right)' = -\frac{f'(x)}{f(x)^2}$, we get

$$= \frac{1}{2}\left((u - g(\Phi))^2 \frac{\frac{\delta}{\delta\Sigma_u}\Sigma_u}{\Sigma_u^2} - \frac{1}{\Sigma_u}\right)$$

$$= \frac{1}{2}\left(\frac{(u - g(\Phi))^2}{\Sigma_u^2} - \frac{1}{\Sigma_u}\right)$$

$$= \frac{1}{2}\left(\frac{(u - g(\Phi))^2}{\Sigma_u^2} - \frac{1}{\Sigma_u}\right)$$

yielding all three equations 2.25.

A.3. Equation 2.28

For moving from the scalar to the matrix form of a predictive network, as done in Equation 2.28, we increase the dimensionality of our toy model organism, which now has observed sensory input $\bar{u}$ and tries to estimate the most likely values $\bar{\Phi}$ of the variables $\bar{v}$. As before, this model has prior expectations that $\bar{v}$ comes from the multivariate normal distribution with mean $\bar{v}_p$ and covariance matrix $\Sigma_{\mathbf{p}}$. Thus:

$$f(\bar{x}, \bar{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^N |\Sigma|}} exp\left[-\frac{1}{2}(\bar{x} - \bar{\mu})^T \Sigma^{-1}(\bar{x} - \bar{\mu})\right]$$

where N is the length of the vector $\bar{x}$ and $|\Sigma|$ is the determinant of the matrix Σ. Hence we have:

$$p(\bar{u}|\bar{v}) = f(\bar{u}; g(\bar{v}, \Theta), \Sigma_{\boldsymbol{u}})$$

where Θ are the parameters of the function g.

Now we can write down the free energy F as:

$$F = \ln p(\bar{\phi}) + \ln p(\bar{u}|\bar{\phi})$$

$$= \ln\left[\frac{1}{\sqrt{(2\pi)^N |\Sigma_{\boldsymbol{p}}|}} exp\left[-\frac{1}{2}(\bar{\phi} - \bar{v}_p)^T \Sigma_{\boldsymbol{p}}^{-1}(\bar{\phi} - \bar{v}_p)\right]\right]$$

$$+ \ln\left[\frac{1}{\sqrt{(2\pi)^N |\Sigma_{\boldsymbol{u}}|}} exp\left[-\frac{1}{2}(\bar{u} - g(\bar{\phi}, \Theta))^T \Sigma_{\boldsymbol{u}}^{-1}(\bar{u} - g(\bar{\phi}, \Theta))\right]\right]$$

$\frac{1}{\sqrt{(2\pi)^N}}$ is a constant, which we group under a constant C term:

$$F = \frac{1}{2}\left[\ln\left(\frac{1}{|\Sigma_{\boldsymbol{p}}|} exp\left(-(\bar{\phi} - \bar{v}_p)^T \Sigma_{\boldsymbol{p}}^{-1}(\bar{\phi} - \bar{v}_p)\right)\right)\right]$$

$$+ \frac{1}{2}\left[\ln\left(\frac{1}{|\Sigma_{\boldsymbol{u}}|} exp\left(-\bar{u} - g(\bar{\phi}, \Theta))^T \Sigma_{\boldsymbol{u}}^{-1}(\bar{u} - g(\bar{\phi}, \Theta))\right)\right)\right] + C$$

$$= \frac{1}{2}\left[\ln\left(\frac{1}{|\Sigma_{\boldsymbol{p}}|}\right) + \ln\left(exp\left(-(\bar{\phi} - \bar{v}_p)^T \Sigma_{\boldsymbol{p}}^{-1}(\bar{\phi} - \bar{v}_p)\right)\right)\right]$$

$$+ \frac{1}{2}\left[\ln\left(\frac{1}{|\Sigma_{\boldsymbol{u}}|}\right) + \ln\left(exp\left(-\bar{u} - g(\bar{\phi}, \Theta))^T \Sigma_{\boldsymbol{u}}^{-1}(\bar{u} - g(\bar{\phi}, \Theta))\right)\right)\right] + C$$

$$= \frac{1}{2}\left[-\ln(|\Sigma_{\boldsymbol{p}}|) - (\bar{\phi} - \bar{v}_p)^T \Sigma_{\boldsymbol{p}}^{-1}(\bar{\phi} - \bar{v}_p) - \ln(-|\Sigma_{\boldsymbol{u}}|) - (\bar{u} - g(\bar{\phi}, \Theta))^T \Sigma_{\boldsymbol{u}}^{-1}(\bar{u} - g(\bar{\phi}, \Theta))\right]$$
$$+ C$$

As we will now express everything in matrix term, it is useful to remember some

properties, such as the gradient on vectors:

$$\bar{x} = \begin{bmatrix} x_1 \\ x_1 \end{bmatrix}$$

if $y = \bar{x}^T \bar{x} = x_1^2 + x_2^2$ by construction, so the gradient becomes

$$\frac{\delta y}{\delta \bar{x}} = \begin{bmatrix} \frac{\delta y}{\delta x_1} \\ \frac{\delta y}{\delta x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} = 2\bar{x}$$

Knowing that Σ matrices are symmetric (because they are covariance matrices), we can now compute the gradient of:

$$F = \frac{1}{2}\left[-\ln\big(|\Sigma_p|\big) - (\bar{\phi} - \bar{v}_p)^T \Sigma_p^{-1}(\bar{\phi} - \bar{v}_p) - \ln(-|\Sigma_u|) - (\bar{u} - g(\bar{\phi}, \Theta))^T \Sigma_u^{-1}(\bar{u} - g(\bar{\phi}, \Theta))\right] + C$$

which is:

$$\frac{\delta F}{\delta \bar{\phi}} = \frac{1}{2}\left[-\frac{\delta F}{\delta \bar{\phi}}\ln\big(|\Sigma_p|\big) - \frac{\delta F}{\delta \bar{\phi}}(\bar{\phi} - \bar{v}_p)^T \Sigma_p^{-1}(\bar{\phi} - \bar{v}_p)\right]$$

$$-\frac{1}{2}\left[\frac{\delta F}{\delta \bar{\phi}}\ln(-|\Sigma_u|) - \frac{\delta F}{\delta \bar{\phi}}(\bar{u} - g(\bar{\phi}, \Theta))^T \Sigma_u^{-1}(\bar{u} - g(\bar{\phi}, \Theta))\right]$$

$$+\frac{\delta F}{\delta \bar{\Phi}}C$$

$$= \frac{1}{2}\left[-\frac{\delta F}{\delta \bar{\phi}}(\bar{\phi} - \bar{v}_p)^T \Sigma_p^{-1}(\bar{\phi} - \bar{v}_p) + \frac{\delta F}{\delta \bar{\phi}}(\bar{u} - g(\bar{\phi}, \Theta))^T \Sigma_u^{-1}(\bar{u} - g(\bar{\phi}, \Theta))\right]$$

We can use the rule that $\delta a x^2/\delta x = 2ax$ to change this form into:

$$\frac{\delta F}{\delta \bar{\phi}} = \frac{1}{2}\left[-2\Sigma_p^{-1}(\bar{\phi} - \bar{v}_p) + \frac{\delta F}{\delta \bar{\phi}}(\bar{u} - g(\bar{\phi}, \Theta))^T \Sigma_u^{-1}(\bar{u} - g(\bar{\phi}, \Theta))\right]$$

and same for the right term for second rule where $z = F, \bar{x} = \bar{\phi}, \bar{y} = \bar{g}$:

$$\frac{\delta F}{\delta \bar{\phi}} = \frac{1}{2}\left[-2\Sigma_p^{-1}(\bar{\phi} - \bar{v}_p) + 2\frac{\delta g(\bar{\phi}, \Theta)^T}{\delta \bar{\phi}}\Sigma_u^{-1}(\bar{u} - g(\bar{\phi}, \Theta))\right]$$

$$\frac{\delta F}{\delta \bar{\phi}} = -\Sigma_p^{-1}(\bar{\phi} - \bar{v}_p) + \frac{\delta g(\bar{\phi}, \Theta)^T}{\delta \bar{\phi}}\Sigma_u^{-1}(\bar{u} - g(\bar{\phi}, \Theta))$$

As previously, we can change terms so the prediction errors simplify the expression:

$$\bar{\epsilon}_p = \Sigma_p^{-1}(\bar{\phi} - \bar{v}_p)$$

$$\bar{\epsilon}_u = \Sigma_u^{-1}(\bar{u} - g(\bar{\phi}, \Theta)$$

Then the gradient becomes

$$\dot{\phi} = -\bar{\epsilon}_p + \frac{\delta g(\bar{\phi}, \Theta)^T}{\delta \bar{\phi}} \bar{\epsilon}_u$$

Note that $\frac{\delta g(\bar{\phi}, \Theta)^T}{\delta \bar{\phi}}$ is a matrix that contains the partial derivative of the element i of $g(\bar{\phi}, \Theta)$ over ϕ_j, i.e. each element is the derivative with a specific parameter theta. For a 2D stimulation, we can then write this as:

$$\frac{\delta g(\bar{\phi}, \Theta)}{\delta \bar{\phi}} = \begin{bmatrix} \theta_{1,1} h'(\phi_1) & \theta_{1,2} h'(\phi_2) \\ \theta_{2,1} h'(\phi_1) & \theta_{2,2} h'(\phi_2) \end{bmatrix}$$

so

$$\dot{\phi} = -\bar{\epsilon}_p + \frac{\delta g(\bar{\phi}, \Theta)^T}{\delta \bar{\phi}} \bar{\epsilon}_u = -\bar{\epsilon}_p + h'(\bar{\phi}) \times \Theta^T \bar{\epsilon}_u$$

where $\times$ is a element-wise multiplication. The gradient on the nodes become

$$\dot{\epsilon}_p = \bar{\phi} - \bar{v}_p - \Sigma_p \bar{\epsilon}_p$$

$$\dot{\epsilon}_u = \bar{u} - \Theta h(\bar{\phi}) - \Sigma_u \bar{\epsilon}_u$$

once more, as done in the previous section of this Appendix, one can derive for parameters $\bar{v}_p, \Sigma_p, \Sigma_u$ to find the expressions:

$$\frac{\delta F}{\delta \bar{v}_p} = \bar{\epsilon}_p$$

$$\frac{\delta F}{\delta \Sigma_p} = \frac{1}{2}(\bar{\epsilon}_p \bar{\epsilon}_p{}^T - \Sigma_p^{-1})$$

$$\frac{\delta F}{\delta \Sigma_u} = \frac{1}{2}(\bar{\epsilon}_u \bar{\epsilon}_u{}^T - \Sigma_u^{-1})$$

which are the expression given in Equation 2.28.

The logic behind the addition of an inhibitory neuron to make computation Hebbian again is the following. A single prediction error must converge to:

$$\epsilon_p = \frac{\Phi - v_p}{\Sigma_p}$$

Where the mean expected level of a feature Φ varies with Σ_p:

$$\Sigma_p = \langle (\Phi - v_p)^2 \rangle$$

Adding an inhibitory node or neuron yields:

$$\dot{\epsilon}_p = \bar{u} - g(\bar{\phi}) - e_p$$
$$\dot{e}_p = \Sigma_p \epsilon_p - e_p$$

By setting the desired value to 0, we get a fixed point at the desired value:

$$\varepsilon_p = \frac{\Phi - v_p}{\Sigma_p}$$

$$e_p = \Phi - v_p$$

For matrix form, the idea is the same as in the previous section of the appendix, where we work with a variance matrix instead of a scalar value:

$$\dot{\bar{\varepsilon}}_p = \bar{\phi}_p - g(\bar{\phi}_{i+1}) - \bar{e}_p$$

$$\dot{\bar{e}}_p = \Sigma_{\mathbf{p}}\bar{\varepsilon}_p - \bar{e}_p$$

As before, we can find the fixed point by setting these variables to 0:

$$\bar{\varepsilon}_p = \Sigma_{\mathbf{p}}^{-1}\bar{\phi}_p - g_p(\bar{\phi}_{i+1})$$

$$\bar{e}_p = \bar{\phi}_p - g_p(\bar{\phi}_{i+1})$$

Thus we can see that nodes ε have fixed points at the values equal to the prediction errors. We can now consider a learning rule analogous to that in the previous subsection:

$$\Delta\Sigma_p = \alpha(\bar{\varepsilon}_p \bar{e}_p^T - 1).$$

To find the values to vicinity of which the above rule may converge, we can find the value of Σ_p for which the expected value of the right-hand side of the above equation is equal to 0:

$$\langle \bar{\varepsilon}_p \bar{e}_p^T - 1 \rangle = 0.$$

B. Appendix B: Pulvinar and Predictive Coding Review

"Daddy sang bass
Mama sang tenor,
CMe and little brother would join right in there."
Johnny Cash, Daddy Sang Bass, 1969

Full citation is as follows: Nelson Cortes, Hugo J Ladret, Reza Abbas-Farishta, et al. "The pulvinar as a hub of visual processing and cortical integration". In: *Trends in Neurosciences* (2023)

CellPress

Review

The pulvinar as a hub of visual processing and cortical integration

Nelson Cortes,[1] Hugo J. Ladret,[1,2] Reza Abbas-Farishta,[1] and Christian Casanova [1,*]

The pulvinar nucleus of the thalamus is a crucial component of the visual system and plays significant roles in sensory processing and cognitive integration. The pulvinar's extensive connectivity with cortical regions allows for bidirectional communication, contributing to the integration of sensory information across the visual hierarchy. Recent findings underscore the pulvinar's involvement in attentional modulation, feature binding, and predictive coding. In this review, we highlight recent advances in clarifying the pulvinar's circuitry and function. We discuss the contributions of the pulvinar to signal modulation across the global cortical network and place these findings within theoretical frameworks of cortical processing, particularly the global neuronal workspace (GNW) theory and predictive coding.

The transthalamic visual cortico-cortical pathways

When viewing an image, signals originating from the retina are transmitted through the lateral geniculate nucleus (LGN) of the thalamus to the primary visual cortex (V1) for local feature processing. Subsequently, information regarding the image characteristics is conveyed to hierarchically organized cortical areas, where progressively greater integration of biologically relevant patterns leads to appropriate behavioral responses [1]. Historically, visual perception was attributed solely to hierarchical cortico-cortical connections [2]. This corticocentric viewpoint has often overlooked the fact that the operations of the cortex are deeply intertwined with those of subcortical structures, including the basal ganglia, cerebellum, and thalamus. This complex interplay forms the basis of our cognitive abilities and behaviors. As depicted in Figure 1, it is important to recognize that sensory information from lower cortical areas can be conveyed directly to higher-order cortical areas not only through cortico-cortical connections but also indirectly via cortico-thalamo-cortical projections, through **higher-order thalamic nuclei** (see Glossary). While the importance of these transthalamic cortical pathways involving higher-order thalamic nuclei has been acknowledged by several researchers, and notable advancements have been made in recent years, numerous questions regarding the involvement of these nuclei in cortical computations remain unanswered [3,4] and, regrettably, theories of sensory and cognitive integration often neglect the role of the transthalamic pathways. Evidence suggests that the pulvinar, a central higher-order thalamic nucleus, which is reciprocally linked to visual cortical areas and also to auditory, somatosensory, associative, and executive cortical regions, assumes a central role in regulating neuronal signal processing within and between cortices [5,6]. Further supporting this proposition, dysfunctions in transthalamic cortical communication via the pulvinar are linked to sensory deficits observed in conditions like autism, attention-deficit/hyperactivity disorder, and schizophrenia [7–9].

Reframing the pulvinar's role within theories that elucidate the integration and computation of diverse cortical signals is essential for a more comprehensive understanding of transthalamic pathways. This review begins with an overview of the visual pulvinar and its associated cortical network. We then delve into the pulvinar's key role in governing visual cortical processing,

Highlights

The thalamic pulvinar exhibits extensive reciprocal connections with the visual cortex. These connections provide alternative pathways for the transmission of signals between cortical neurons.

Although the precise roles of the pulvinar in vision are incompletely understood, research indicates that the pulvinar is involved in target selection among distractors, feature binding, and visual attention. Pulvinar dysfunctions have been associated with sensory deficits observed in individuals with schizophrenia.

Computational models suggest that the pulvinar plays a crucial role in establishing persistent activity within the cortical network by regulating the flow of information across visual cortical areas and the synchronization of their activity.

We highlight the organization of brain networks involving the pulvinar and propose a pivotal role for the pulvinar within the framework of cortical theories that aim to elucidate the processes of perception and cognitive integration.

[1]Visual Neuroscience Laboratory, School of Optometry, Université de Montréal, Montreal, QC, Canada
[2]Institut de Neurosciences de la Timone, UMR 7289, CNRS and Aix-Marseille Université, Marseille, 13005, France

*Correspondence:
christian.casanova@umontreal.ca
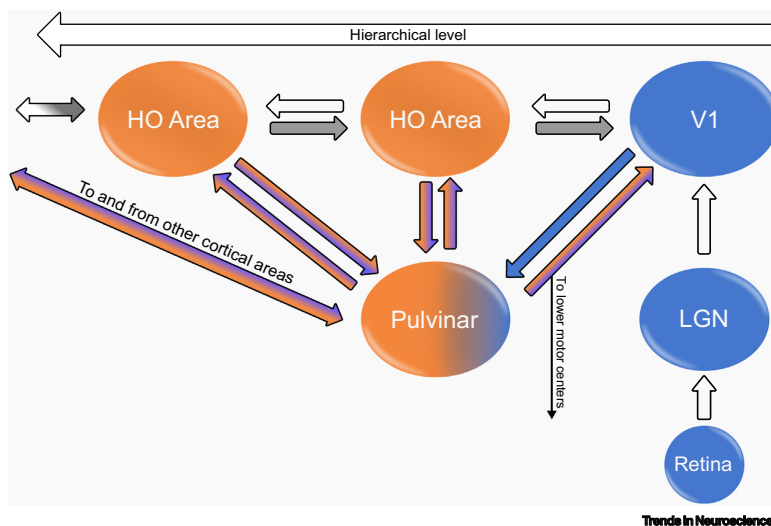(C. Casanova).

Trends in Neurosciences

CellPress

Figure 1. The transthalamic cortical visual pathways. Sensory signals from the retina are sent to the lateral geniculate nucleus (LGN) and then to the primary visual cortex (V1). V1 signals can be transmitted to higher-order visual cortical areas (HO) not only through cortico-cortical connections (white and gray arrows, representing feedforward and feedback projections, respectively) but also through the pulvinar, which mediates the transthalamic cortical communication between all areas of the visual cortex in the feedforward and feedback directions.

specifically its integration of visual and cognitive information across cortical areas. We pay special attention to cortical theories that aim to clarify perception and cognitive integration, such as the GNW theory and predictive coding frameworks.

Subdivisions and functional connectivity of the pulvinar

The pulvinar is situated above the dorsolateral posterior thalamus and extends along the medial border of the LGN. Throughout mammalian evolution, the pulvinar has grown and differentiated, mirroring changes in the neocortex (Box 1) [10]. Consequently, the pulvinar stands as the most prominent visual nucleus within the thalamus in higher-order mammals, including humans [11,12]. The subsequent section delineates the anatomical subdivisions of the pulvinar in primates, cats, and rodents, emphasizing its multifaceted roles from sensory processing to motor control. Next, we introduce the concept of a dorso-ventral pulvinar gradient as a framework to better understand the functional diversity of the pulvinar.

Anatomical subdivisions in primates

Historically, in view of the pulvinar's intricate architecture in primates, this brain region was divided into three major subdivisions: the inferior pulvinar (PI), the lateral pulvinar (PL), and the medial pulvinar (PM) [13]. A contemporary naming system has refined pulvinar subdivisions based on its connection with early visual areas (Box 1) [14]. Thus, the PI and PL have been segmented further. The PI is subdivided into the posterior, middle, central medial, and central lateral subnuclei (PIp, PIm, PIcm, and PIcl, respectively). While PIp, PIm, and PIcm primarily interface with the **dorsal visual stream**, PIcl exhibits connections with the **ventral visual stream** [15]. The PL has been further categorized into two regions: the dorso-medial (PLdm) and the ventro-lateral (PLvl). While PLvl is primarily linked to the superior colliculus (SC), V1, V2, and areas associated

Glossary

Alpha rhythm: rhythmic neural activity typically defined in humans within the 7.5–12.5 Hz range.
Asynchronous or synchronous neuronal activity: asynchronous activity in neural networks describes a regime with weak temporal correlations among neurons, whereas synchronous activity refers to a regime where temporal correlations are strong.
Beta rhythm: rhythmic neural activity typically defined in humans within the 12.5–30 Hz range.
Dorsal visual stream: the dorsal stream, or 'where/how' pathway, originates in the striate cortex, extending to the parietal lobe for spatial and motion awareness.
Feedforward and feedback processing: feedforward processing refers to flow of neural information through hierarchical levels of processing, from low level to higher levels, whereas feedback processing is the reverse flow from higher to lower levels, enabling top-down modulation of activity.
Firing rate coding: in firing rate coding, the intensity of a stimulus is encoded by the frequency or rate of neuronal action potentials. Excitatory neurons typically increase firing rates for stronger stimuli; inhibitory neurons decrease firing rates, modulating neural network activity.
Fleeting memory: also known as sensory memory, it is the initial stage of memory processing that briefly stores incoming sensory information in its original form before it is further processed.
Gamma rhythm: rhythmic neural activity typically defined in humans within the 25–70 Hz range.
Higher-order thalamic nucleus: unlike first-order nuclei, which primarily receive their driving input from the sensory organs, higher-order thalamic nuclei receive their primary driving signals from cortical areas.
Subliminal visual stimulus: typically refers to a brief and low-energy visual stimulation that the observer does not consciously perceive.
Temporal coding: a mode of neuronal communication that relies on the precise timing of action potentials or inter-spike intervals for encoding information.
Theta rhythm: rhythmic neural activity typically defined in humans within the 4–7.5 Hz range.
V4: a cortical visual area of the ventral pathway that integrates feature

**Box 1. Evolutionary expansion of the pulvinar in primates: functional nuances and the role of vGLUTs**

The primate pulvinar exhibits two prime distinctions from other species: the expansion of PM, and the proliferation of the ventral subunits, PI and PLvl (Figure I). PM, as highlighted by recent research, functions as a multisensory hub, crucial for sensory–motor integration, eye–hand coordination, and potentially conveying complex emotional cues to the amygdala [8,16]. In contrast, PI and PL, which have evolved into at least five subdivisions, enhance the relay from sources such as V1 and V2, retina, and SC to both dorsal and ventral visual streams in the cortex [13–15]. For instance, PIcm and PIp primarily interact with regions around the medial temporal area (MT) and receive direct projections from SC, and PIm, receiving extra projections from the retina, has a direct link with MT [12,15,39,116,118]. A potential driving factor behind the evolution of ventral subdivisions might be the migration of neurons from the SC to the retinofugal pathway. For context, 85–90% of rodent retinal cells project to the SC [119], but only about 50% in cats (26% from temporal retina) [120] and 10% in primates do so [121]. Such migrations illustrate the balance between retinal and SC projections into the caudal pulvinar, influencing the evolution of the PI and the emergence of the MT complex [118,122]. Collectively, these changes in the pulvinar allow for intricate interactions with the cortex, accommodating high-level neural integration, as seen with PM, and specialized visual perception tasks, as seen with PI and PL.

The newly discovered subdivisions of PL and PI exhibit distinctive protein staining patterns and varied connections, revealing diverse interactions with visual cortical areas and a range of functions [14]. A significant example of this complexity is the divergent roles of vesicular glutamate transporters, vGLUT1 and vGLUT2, essential markers for delineating functionally distinct territories within the visual pulvinar and for understanding the dynamics of thalamocortical neurons and their inputs. Extensive studies have investigated the unique roles of vGLUT1 and vGLUT2 across diverse species, uncovering their subtle functionalities within the visual pulvinar [123]. Initially, vGLUT1 was associated with modulatory projections and vGLUT2 with driving projections [124]. However, this initial dichotomy has proven to be more nuanced. For instance, in macaques, driver corticopulvinar projections from layer 5 pyramidal cells are enriched with vGLUT1, while vGLUT2 defines subcortical terminal areas [12]. These areas may be crucial for relaying information from the retina and SC to the neocortex. These interspecies variations stress transporters' adaptive roles in visual processing, showing the balance in neural and synaptic functions across diverse contexts.

processing and visual attention, situated at intermediate levels of the visual cortical hierarchy.

Ventral visual stream: the ventral stream, the 'what' pathway, also starts in the striate cortex but leads to the temporal lobe, focusing on object and face recognition.

Winner-take-all: in the context of neural networks, winner-take-all is a competitive neural mechanism where only a specific subset of neurons fire in response to a particular input.
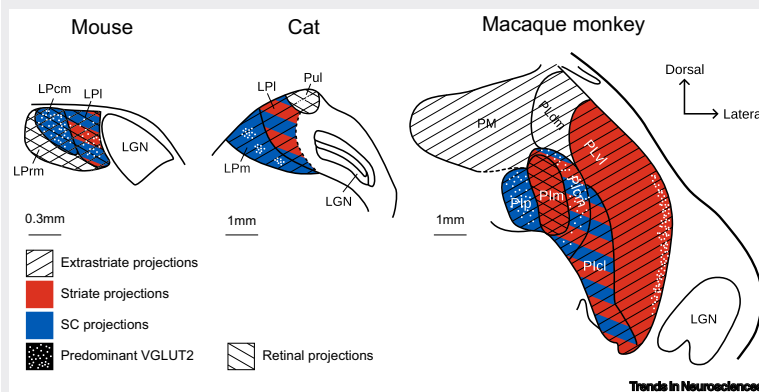
Figure I. This figure illustrates the distinctive pulvinar subdivisions present in mice, cats, and macaques, their VGLUT2 expression, and a general view of their cortical and subcortical inputs. In mice, the subdivisions of the lateral posterior nucleus (LP, homolog of pulvinar) are as follows: LPcm, caudal medial LP; PI, lateral LP; LPrm, rostral medial LP. In cats, the main subdivisions of the LP-pulvinar complex are designated as: LPl, lateral LP; LPm, medial LP; Pul, the pulvinar nuclei proper. In macaques, a more complex structure is observed, including: PIcm, central medial inferior pulvinar; PIcl, central lateral inferior pulvinar; PIp, posterior inferior pulvinar; PIpl, posterior lateral inferior pulvinar; PLvl, ventrolateral lateral pulvinar; PIdm, dorsomedial lateral pulvinar; PM, medial pulvinar. Please note the prominent increase in the most dorsal subunit, PM, and the proliferation of subunits devoted specifically to visual processing in macaques, illustrating the advanced and specialized nature of the primate pulvinar relative to other species. Mice, cat, and macaque schematics are based on [12,23,125], [19,20,22,98,107,124,126,127], and [12–15], respectively. Abbreviations: LGN, lateral geniculate nucleus; SC, superior colliculus.

Trends in Neurosciences

with the ventral visual stream, PLdm is more associated with regions that are part of the dorsal visual stream [16]. Finally, the PM has two specialized subdivisions [5,17]. Its lateral division involves extensive visual processing, connecting with areas such as the LGN, V1, V2, dorsal and ventral stream cortical areas, and the SC. In contrast, PM's medial division is linked to auditory and motor-related regions, signifying a broader range of functions (see later and Box 1) [8].

Comparative perspectives: cats and rodents
In cats, the pulvinar manifests as the lateral-posterior (LP)/pulvinar complex, with the lateral and medial portions of the LP nucleus (LPl and LPm, respectively), and the pulvinar proper [18]. While the LPl has a unique feature of direct projections from the striate cortex or V1 [18,19], both LPl and LPm receive dense recipient extrastriate projections. However, LPm is the main recipient of SC projections, with LPl having a small tectorecipient zone [20,21]. Additionally, the pulvinar proper receives primarily retinal inputs [22]. The relationship between subnuclei and dorsal/ventral cortical streams in cats remains largely unknown. In rodents, the pulvinar is notably smaller and more homogeneous than in higher mammals [23]. Nevertheless, recent research found distinct subdivisions with unique cortical inputs in rodents, implying specialized subregions [24].

Dorso-ventral pulvinar gradient
One of the main challenges in the study of pulvinar physiology is that its functions in sensory processing are sometimes inconsistent with its anatomical subdivisions [25,26]. Despite the complexities that arise from the pulvinar's anatomical subdivisions, the pulvinar exhibits a dorso-ventral functional gradient, which emulates a hierarchical continuum of brain functions from basic perception to advanced cognitive processing [5]. In primates, the ventral pulvinar, often referred to as the 'visual pulvinar', includes the anatomically defined PI and PLvl, both of which are intricately connected to occipital and temporal cortices [16]. Specialized connectivity shapes pulvinar's receptive fields, which bear a striking resemblance to those of visual cortical neurons. Research in both cats and monkeys has shown that many pulvinar receptive fields are binocularly selective and sensitive to retinal disparity [27]. Additionally, they respond to essential image features like orientation and motion direction [20,28]. Further studies revealed that pulvinar neurons in cats and humans also encode higher-order visual processes, such as discerning the direction of complex visual stimuli like moving plaids and random dot kinematograms [29–31]. Moreover, a coarse retinotopic organization has been identified in the lateral and inferior primate pulvinar subdivisions [32,33], as well as in the main subdivisions of the LP-pulvinar in cats [20,34]. Conversely, the dorsal pulvinar of primates, consisting of the PLdm and PM, aligns more with higher-order cognitive processes [16]. It is functionally connected to frontal, parietal, and cingulate cortices and is essential for attentional control (e.g., goal-directed eye movements) and other advanced cognitive functions [14,16,35]. For example, dorsal pulvinar, like PM, processes fear-eliciting stimuli like images of snakes through its connectivity with the amygdala [36,37]. This conserved behavior may underlie phenomena like 'affective blindsight', which relies less on corticothalamic or thalamocortical connectivity and more of a PM-driven event [26,38,39]. In sum, the pulvinar's dorso-ventral functional gradient serves as a unifying framework that encompasses its diverse functions, ranging from basic visual processing to advanced cognitive functions [20,27–32,36,37,40,41].

Unveiling the complexity of cortex–thalamus interactions: an integrated examination of a driver/modulator framework
Anatomical characteristics of axon terminals
It is noteworthy that a neuron in the primary visual cortex is only one thalamic synapse away from neurons in higher-order areas, thanks to the pulvinar. According to the 'driver/modulator' framework [42], cortical inputs are conceptualized to belong to one of two categories: modulators and drivers (type 1 and type 2 terminals) [43,44]. Type 1 terminals, with small round terminals, thin

axons, and long, thin branches, may adjust or modify the ongoing firing of recipient neurons, through metabotropic receptors [45]. Conversely, type 2 terminals are characterized by large round terminals, large caliber axons with clustered endings [18]. They define the target cell's receptive field using fast ionotropic receptors [45,46]. Cortico-thalamic projections to pulvinar neurons exhibit a mix of these two terminal types, but their relative proportion varies across the visual cortical hierarchy [47]. Neurons in early visual cortical areas have a higher proportion of type 2 terminals, while those in higher processing levels send more type 1 terminals [18]. This organizational scheme suggests that early visual areas like V1 send driver signals to form basic visual maps in the pulvinar, whereas higher-order visual cortical areas primarily modulate pulvinar neuron activity. The driver/modulator framework is not immutable. Atypical axon terminals with driver-like properties (type 3) have been discovered in collicular neurons that project to the pulvinar [45]. These may play a vital role in transmitting information about eye fixation and saccades to the cortex [48]. Little is known about the morphology of the reverse connections, the thalamocortical projections from the pulvinar, and their variations across cortical hierarchy. Addressing this knowledge gap would be essential for a comprehensive understanding of the pulvinar-cortical network and its functional implications.

The distinction between drivers and modulators has also been made based on the laminar organization of thalamocortical and corticothalamic projections. For bottom-up projections, the prevailing view suggests that thalamic terminals ending in layer 4 act as drivers, while those ending in layer 1 serve as modulators [49]. In top-down projections, layer 5 cortical cells provide driver signals, whereas layer 6 cells transmit modulatory signals. This holds true for the LGN, which targets layer 4 in V1 and receives feedback modulatory signals from layer 6 [50]. While it is generally accepted that pulvinar neurons in primates and cats receive driver inputs from V1 [51,52], and in turn send modulatory signals to layer 1 [53,54], it is still unknown if corticothalamic pathways involving higher-order areas and the pulvinar conform with the driver/modulator framework, despite recent findings of hierarchical anatomical organization of cortico-pulvinar projections in cats [18]. Addressing this issue is especially challenging since pulvinar axon terminals or cortical cell bodies are generally not restricted to a single cortical layer [55,56].

Functional characteristics of axons terminals

Drivers and modulators can be classified based on the input–output transfer of the neuronal responses. Modulator effects are categorized as either multiplicative or divisive (nonlinear gain control), while drivers act through additive or subtractive changes (linear gain control) [57,58]. Studies in rats have demonstrated that type 1 and type 2 corticothalamic terminals from V1 exhibit short-term facilitation and depression in excitatory postsynaptic potentials, respectively [59]. Expanding on this, the driver/modulator framework can indirectly correspond to the anatomical organization of the cortical visual hierarchy. Such a connection supports a unique type of communication throughout the hierarchy, deeply rooted in the inherent rhythmicity delineated in the local field potential (LFP) of neurons. In primates, feedforward connections are associated with high-frequency rhythms (gamma oscillations, >30 Hz), while feedback connections are linked to low-frequency oscillations (alpha and beta oscillations, 7–30 Hz) [60–62].

In cats and monkeys, findings from neuronal inactivation studies align with corticothalamic anatomical features. Silencing V1, for instance, abolishes visual responses in most pulvinar neurons in cats and monkeys [51,52], underscoring the notion of V1 as a driver of pulvinar activity. Conversely, considering our previous anatomical discussion, higher-order cortical areas are believed to primarily modulate pulvinar activity. However, further investigation is required since, in cats, higher-order areas are necessary to drive pattern-motion neurons in the pulvinar [30]. Similarly, limited information is available about the pulvinar's impact on cortical neurons in the visual hierarchy.

Trends in Neurosciences

CellPress

In primates, deactivating the PL notably reduces activity in V1's supragranular layers [54]. In cats, the primary effect of LPl inactivation on V1 neurons is a slight decrease in response gain [53], corroborating previous findings obtained through optical brain imaging in the same species [63]. In contrast, LPl inactivation leads to a substantial increase in response gain in most neurons in area 21a (putative homolog of area **V4** in primates), with a few cells exhibiting contrast gain changes [53]. Similarly, in *Cebus* monkeys, PL inactivation intensifies response amplitude in area V2 [64]. Collectively, these studies, while relatively limited, confirm that the pulvinar can influence processing across the visual cortex mainly by modulating neuronal activity, thus suggesting its role in controlling the flow of information throughout the visual cortex.

The pulvinar's dual role in visual attention and neural communication
Understanding the precise role of the pulvinar is a challenging task due to its unique anatomical-functional architecture. This thalamic region has been implicated in various processes and readers are encouraged to refer to previous reviews that describe the potential functions associated with the pulvinar [3,65–67]. In the next paragraphs, we will examine two prominent roles debated in the literature, relevant to the subsequent discussion on theoretical models.

Modulating visual attention
For decades, the pulvinar has been recognized as a modulator of visual attention. Early studies in monkeys suggested that neurons involved in attention tasks were primarily found in the dorsal pulvinar (i.e., in PLdm) [68,69]. Deactivating PLdm shifted visual attention and affected the motivation to reach and grasp objects, mirroring perceptual neglect symptoms. Subsequent investigations utilizing brain imaging and psychophysical techniques in both healthy and brain-damaged individuals reinforced the pulvinar's role in sustained visual attention, stimulus awareness, and the filtering of irrelevant or distracting information [23,70–77]. Likewise, a recent study in humans showed that activity in the pulvinar and fronto-parietal cortices was modulated when a cue and a salient distractor were presented together [78]. However, this modulation did not occur when either of these stimuli was presented alone. Relatedly, an fMRI study in humans revealed increased activity in the dorsal pulvinar during directed spatial attention, mirroring fronto-parietal activation patterns [25]. Single-cell recordings in the lateral and inferior regions of the ventral pulvinar in primates [79,80] and in the LP-pulvinar complex of ferrets [81] align with the findings from humans. These animal studies revealed that pulvinar neurons are involved in sustained and directed spatial attention, although the specific neuronal activity (e.g., LFP rhythms such as alpha, theta, or gamma oscillations) associated with attention varies across these investigations.

Orchestrating cortical communication in neural networks
Another important function of the pulvinar, besides its role in visual attention, is its capacity to modulate information transmission within large-scale cortical networks. Two types of cortical processing, **feedforward and feedback processing**, underlie whole-brain activation. Extensive theoretical research proposes pulvinar-mediated feedforward cortical amplifications, backed by empirical evidence. Computational models, for instance, indicate that the pulvinar enhances external input responses throughout the cortical hierarchy, acting as a **firing rate coding** mechanism in visual processing [82–84]. Studies conducted in cats provide further support for this feedforward amplification, as, counterintuitively, inactivation of the lateral and medial subdivisions of the LP/pulvinar complex leads to increased spike counts tied to contrast changes in cortical areas at various hierarchical positions [53]. Similarly, mouse brain slice research indicates that the pulvinar's activity may exhibit driver-like characteristics, amplifying signals through transthalamic feedforward pathways. Notably, in these experiments, most of the pulvinar projections, originating from zones that receive inputs from the striate cortex and the SC, target layer 4 of extrastriate cortices [85]. These findings underscore the critical role of the pulvinar in visual

processing by highlighting its involvement in the contextual integration of cortical information, ultimately influencing higher cortical areas.

Expanding on these insights, the pulvinar's role in coordinating information transmission in cortical networks becomes clearer when considering **temporal coding**. In the context of feedforward propagation, transthalamic pathways are organized in such a way that they facilitate the emergence of nested oscillatory patterns within different cortical subnetworks [80,86]. Recent research supports this organization by revealing synchronized LFP coherence between cortical regions and the PLdm during both ongoing activity and attention-dependent tasks [80,87,88]. Notably, recent data suggest a direct link between pulvinar modulation and the amplification of **gamma rhythms** [86]. Conversely, neuronal populations in the prefrontal cortex generate oscillatory states characterized by low-frequency oscillations that align with pulvinar rhythms, including **theta rhythms**, **alpha rhythms**, and **beta rhythms**, primarily within the PM and PLdm [73,88]. Therefore, the pulvinar can be seen as a dynamic trigger facilitating communication between distinct cortical areas through the synchronization of oscillatory patterns, phase relationships, and rhythmic activity.

While the role of the pulvinar in feedback transmission is not yet fully understood, theoretical studies suggest that transthalamic pathways transmit alpha and beta rhythms back to the cortex [84]. Experimental data indicate that similar feedback-related low-frequency signatures observed in cortical areas of monkeys [60,62] also manifest in the human pulvinar [89]. These rhythms are predominantly generated within the cortex, particularly the prefrontal cortex [79,88,90]. However, eliciting thalamic responses requires significant effective connectivity, which can only occur if multiple cortical areas that project to the pulvinar are synchronously activated (given the modulatory nature of type 1 terminals). The precise number of top-down inputs (i.e., corticothalamic inputs) needed to trigger pulvinar responses remains unclear. An alternative hypothesis suggests that cortical inputs associated with the thalamic reticular nucleus generate a global inhibitory response within the pulvinar, leading to low-frequency oscillations [91]. Another mechanism suggests that slow oscillations may arise from asynchronous corticothalamic projections into discrete pulvinar clusters, leading to **winner-take-all** states of neuronal competition [47]. This hypothetical neuronal mechanism could facilitate the resetting of cortical feedback transmission by generating a bistable state (either **asynchronous or synchronous neuronal activity**) within the pulvinar, which is influenced by the distinct distribution of corticothalamic type 1 and 2 axons in the visual cortex. A similar bistable state has been experimentally observed in the primate dorsal and ventral pulvinar during object detection and passive viewing [92].

The preceding paragraphs have highlighted the pivotal position of the pulvinar within the global neural network of the brain. Moving forward, we will examine prevailing brain theories and explore how the pulvinar integrates into these concepts, bridging the gap between current understanding and future research directions.

The global workspace theory (GWT) and GNW

The GWT is a prominent cognitive science theory that explains how distributed modules communicate to achieve higher-order cognitive states, including consciousness [93]. In this theory, a **fleeting memory** enables the system to access shared brain functions, and long-range connections between modules facilitate large-scale network operations [94] The GWT posits that sensory brain structures provide network inputs, with modules filtering out irrelevant information to share relevant activity. Consequently, each module collaborates as a functional part of the system, unifying global meaning at each moment. An extension of the GWT, the GNW model, offers further insights into the neural architecture and workings of the brain in generating

Trends in Neurosciences

CellPress

conscious experiences [95,96]. An example of the GNW in action is the neuronal processing during the conscious detection of a **subliminal visual stimulus** [97]. This detection process unfolds in several steps within the cortex: crossing a cortical threshold, feedforward broadcasting, reverberating cortical activity, and feedback neuronal communication. The dynamics of this process, called 'ignition' in the GNW, can be summarized as follows [98]. When a signal from a visual stimulus reaches V1, it crosses a neural threshold, triggering activity propagation throughout the cortex. This activity cascades from lower to higher hierarchical levels, broadcasting feedforward information up to the prefrontal cortex. However, if the signal is not sufficiently strong, it decays exponentially throughout the hierarchy (Figure 2A$_1$). If the threshold is reached and the activity spreads across the entire cortex, the neuronal activity continues to reverberate through recurrent interactions within and between cortical areas [97]. The global neuronal response is then relayed back through the cortex toward lower-level regions. This process involves bidirectional communication, with feedforward and feedback connections both being essential to conscious experience. Cortico-cortical projections, especially long-range connections, contribute to nervous system integration and greatly impact a subject's conscious states from the GNW perspective.

Integrating the pulvinar in the GNW

Based on findings discussed in earlier sections, the pulvinar may assume a critical role within the GNW by contributing to one or multiple stages of cortical ignition. These stages encompass the activation threshold, feedforward propagation, recurrent activation, and feedback integration processes (as depicted in Figure 2A$_2$).

As highlighted earlier, the pulvinar likely enhances cortico-cortical communication by amplifying signals through the feedforward pathway. This amplification is achieved through the pulvinar's extensive projections to extrastriate cortical areas, particularly targeting layer 4. We propose that this transthalamic pathway serves to enhance the feedforward transfer of information between cortical modules by increasing the response gain along the cortical hierarchy, minimizing the dampening of signal propagation over distance [82,83,99]. Additionally, the pulvinar might be involved in cortical feedback processes, potentially facilitated by its hypothetical bistable activation or global inhibition associated with the thalamic reticular nucleus [91]. The cortico-pulvino-cortical feedback action may serve the purpose of resetting cortical feedback transmission, thereby promoting the initiation of subsequent cortical ignitions. In essence, the thalamus can be viewed as a 'controller' of information flow in the brain, regulating the exchange of information among cortical regions to facilitate the establishment of a global workspace.

Is the pulvinar a key player in triggering cortical ignition? There is physiological evidence supporting this notion, particularly in terms of threshold initiation, as electrical microstimulation of the thalamus has shown that the pulvinar can evoke synchronized activity across multiple cortical areas [63,100]. Furthermore, anatomical investigations have revealed the pulvinar's extensive projections to various visual regions as well as multisensory, associative, and executive cortices, indicating its potential to synchronize and initiate cortical activity simultaneously [101,102]. Notably, the burst-like discharges observed in the pulvinar further suggest that such activity enhances the likelihood of activating cortical neurons, allowing them to detect changes in the environment [103].

While the exact 'message' conveyed by the pulvinar to the cortex remains debated, studies in cats and monkeys suggest the pulvinar acts as an information router, transmitting topographical signals to ensure coordinated processing of similar information across distant areas [65]. However, recent findings in mice propose that the information transmitted through the cortico-cortical and transthalamic pathways may exhibit differences [104], indicating that the pulvinar might also play a role in information processing. Nevertheless, it is likely that the transthalamic

216

pathway via the pulvinar contributes to the activation of cortico-cortical connections in the visual cortex, potentially encompassing attentional signals necessary for cortical ignition.

In this theoretical framework, the GNW of the brain is responsible for integrating and disseminating sensory information throughout the cortex. Here, the attentional system acts as a crucial filter that amplifies the most relevant information for conscious experience. Notably, signals from the prefrontal cortex, relayed by nuclei like the pulvinar, play a key role in this attentional filtering [88]. Recent findings underscore that pulvinar's ventral and dorsal regions in humans functionally complement the cortex, activating cortical zones responsible for object visual perception (processed in visual areas) and attention-guided vision (processed in frontal, parietal, and cingulate cortices), respectively [5,25,105]. Thus, given its direct communication with the prefrontal cortex and the SC, both pivotal in attention processes, the pulvinar significantly influences where subjects' attention is directed [11,16,79]. By selectively enhancing and amplifying sensory information, the thalamus may assume an essential role in shaping conscious perception and contributing to the overall global workspace. This role becomes particularly crucial when attention-directing signals from the pulvino-cortical coupling are weak or absent. In such cases, a subliminal stimulus quickly diminished across cortical stages [83], essentially 'turning off' the ignition [97]. In a nutshell, the pulvinar acts as a gatekeeper for thalamocortical coupling, allowing only the most relevant information to enter conscious experience.

The GWT suggests that conscious perception employs fleeting memory to handle limited capacity data, like working memory mechanisms. This involves sustained neural firing rates, even without external stimuli. This activity is observed in primates during figure-ground segregation tasks [106]. While pulvinar neurons' persistent activity remains unknown, two experimental studies found evoked sustained responses in the cat's pulvinar neurons by a visual flash, lasting up to 1 min post-stimulus presentation [107,108]. Recent theories predict sustained pulvinar responses [84], but it is unclear if their activity originates from reciprocal interactions with the cortex or recurrent pulvinar neuron processes. Regardless of the sources, persistent pulvinar responses may drive sustained activity in visual cortical neurons.

Expanding on prior research, we propose an enhanced ignition mechanism involving the pulvinar (Figure 2B). Initially, V1 forwards activity to the pulvinar and adjacent visual areas (Figure 2B_1,2B_2). Once the pulvinar exceeds its threshold (Figure 2B_3), it activates other cortices through transthalamic pathways. If pulvinar activation coincides with cortical feedforward input, cortical activity rapidly expands, enhancing neuronal activity (Figure 2B_4). When multiple higher-order areas target the pulvinar, its activity patterns undergo alterations, leading the pulvinar to disengage from contributing to the ignition (Figure 2B_5). This shift reduces the triggered ignition, allowing for the initiation of a new cycle of GNW (Figure 2B_6). This cyclic ignition and dampening process dynamically regulates higher-order processes.

GNW, predictive coding, and the pulvinar
Another potential explanation for the involvement of the pulvinar in cortical processing is provided by the predictive coding framework, as detailed in Box 2. In many of the initial formulations of predictive coding, similar to the GNW, only cortico-cortical connections were considered in the context of transmitting predictions and prediction errors through feedforward and feedback cortical pathways [109]. Interestingly, early on, a similar theory was suggested where the pulvinar, with its unique thalamic nucleus characteristics, adjusts cortical processes based on a global context of the external world [110]. These frameworks have since evolved to incorporate subcortical structures into their conceptualization of brain function, acknowledging the importance of neuromodulatory mechanisms and transthalamic pathways in neural processing (Figure 3).

Trends in Neurosciences

CellPress
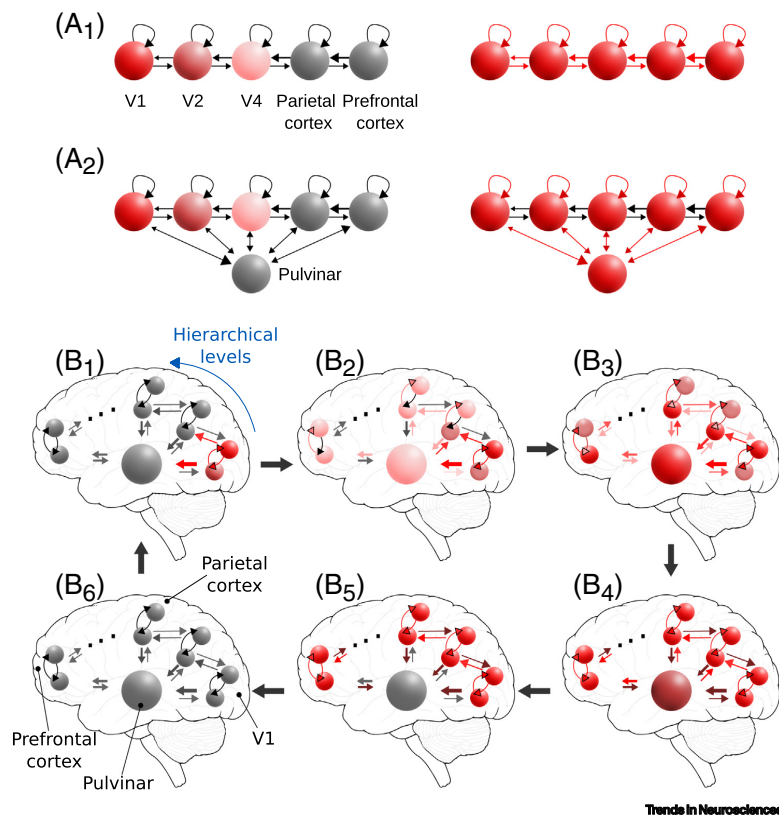
Figure 2. Processing in the global neuronal workspace. (A₁) Upon reaching the primary visual cortex (V1), a visual signal may surpass a neural threshold, enabling activity to propagate across the cortex in a feedforward manner. Without strong stimulation, this signal decays exponentially (left). As cortical areas reach the threshold, activity spreads across the cortex (right), reverberating through intra- and inter-area interactions and returning to lower-level regions. (A₂) Hypothetical contribution of the pulvinar in ignition evolution. As the signal arrives at the pulvinar via V1, activity rapidly expands through cortical areas, amplifying cortical activity if it coincides with feedforward input. (B) Hypothetical global neuronal workspace (GNW) cycle with the pulvinar: (B₁) V1 sends activity to the pulvinar and nearby visual areas. (B₂) The pulvinar activates other cortices via transthalamic cortical pathways, boosting activity (B₃). If targeted by multiple higher areas, pulvinar activity patterns change (B₄), causing the pulvinar to stop contributing to ignition (B₅), dampening the evoked ignition and allowing a new cycle to start (B₆).

This historical perspective aligns with newer theories that incorporate the pulvinar in predictive coding, emphasizing the pulvinar's distinctive role [111]. Beyond processing predictions and prediction errors, the pulvinar may compute the inverse variance (i.e., the precision of signals), thereby adjusting the balance between reliance on prior predictions and integration of new prediction errors [112]. This 'precision-weighting' mechanism may play a pivotal role in perception and cognition, offering a differentiated contribution to cortical processing distinct from prediction handling [113]. According to some conceptualizations, disrupting this mechanism may contribute to psychosis, such as seen in schizophrenia [112,114] (see also Box 3 for further details).

218

⟲ CellPress

Trends in Neurosciences

Box 2. Predictive coding and the pulvinar

Predictive coding is a theory conceptualizing the brain as an organ of inference. According to this framework, the brain builds a model of its environment to craft plausible explanations of the sensory input impinging on it and makes constant predictions about possible future states of the world [128]. The internal model is continuously updated based on prediction errors, that is, mismatches between predictions and actual observed events. Neuroanatomically, this computational principle relies on feedback connectivity carrying (top-down) predictions about the environment, which are matched to feedforward sensory inputs to convey (bottom-up) prediction errors. To regulate the transfer of information between the feedback and feedforward pathways, the brain needs to select which of its representations (predictions), or its new sensory inputs (prediction errors) are best suited to steer behavior accurately. According to Bayesian inference, one of the many computations predictive coding can support [129], the brain weighs this information based on its precision. While prediction errors comfortably match the canonical connectivity of the cortex [128–130], the implementation of precision weighting is still up for debate. Functionally, precision weighting can be described as a gain control mechanism regulating the confidence in either prediction or prediction errors [111,131]. Gain control (modulator) connectivity is a prominent feature of higher-order cortico-thalamic synapses that have been thought to allow cortical areas to encode new salient and relevant features of the environment [132]. The pulvinar's engagement in attentional regulation and its output connectivity to superficial cortical layers is an ideal match for a role in precision modulation of the visual hierarchy, which also fits the long timescales of mesoscale integration associated with precision weighting (see Figure 3 in main text) [111,131]. As described in the main text, the properties of the pulvinar support its role as a modulator of prediction errors throughout the visual hierarchy, thereby ascribing precision to the cortical message passing.

Predictive coding relies on recurrent mechanisms, suggesting the brain's capacity for self-regulating activity within its neural network, as incorporated, for instance, in the GNW. In this context, higher-order thalamic nuclei can be seen as integral components of cortico-cortical loops, rapidly integrating information across different hierarchical levels to amplify bottom-up signals in the presence of low-variance input and vice versa. Notably, this process is not limited to the visual hierarchy alone. Both functional and anatomical evidence suggest that inputs from the prefrontal cortex can directly target V1 neurons through pulvinar axons, forming a macroscale recurrent circuit capable of accessing any cortical area [115]. The anterior cingulate cortex, for example, in mice: a prefrontal region linked to behavioral error and visual orientation actions [116,117]. Whether viewed from the perspective of the GNW or predictive coding framework, incorporating the pulvinar into cortico-cortical theories addresses crucial computational challenges, including hierarchical processing, recurrent propagation, and efficient communication within a distributed network of cortical areas.

Concluding remarks and future perspectives

This review synthesizes the pivotal role of the pulvinar in transthalamic cortico-cortical pathways, emphasizing its influence on visual and cognitive brain functions. It challenges traditional cortico-centric perspectives by highlighting the pulvinar's integration with cortical processes, such as inter-areal neural communication and attentional processing. The anatomical and functional findings of the pulvinar described here align with two key theories in neuroscience: the GNW theory and predictive coding. These findings suggest the pulvinar's vital role in cortical activation stages and fine-tuning neural networks.

The review also underscores unresolved questions. Firstly, the anatomical organization of pathways involving the pulvinar requires further characterization for a comprehensive understanding of physiological roles. Advances in optogenetics and connectomes offer promising avenues for exploring transthalamic connectivity. Future exploration involves understanding how the pulvinar nucleus aids in integrating and segregating cortical networks during complex cognitive functions. This avenue of research is not solely about mapping connections but also about decoding the orchestration of pulvinar and cortical networks in real time as subjects participate in ongoing cognitive tasks. Additionally, a critical area of research involves determining the contribution of the pulvinar to the overall efficiency of brain information processing. Examining the pulvinar's structural similarities with artificial neural networks presents a fascinating

Outstanding questions

What role does the pulvinar nucleus play in the integration and segregation of cortical networks observed during the performance of increasingly complex cognitive perceptual tasks? How does the pulvinar's involvement contribute to the overall efficiency of information processing in the brain?

Are the well-described organizational principles of cortico-cortical neural architecture preserved and maintained in transthalamic pathways, specifically concerning the role of the pulvinar nucleus in integrating and segregating information between cortical regions during the execution of complex cognitive perceptual tasks?

Do individuals with schizophrenia exhibit fewer 'ignition events' (i.e., widespread brain activity triggers) than those without schizophrenia? Do ignition events in patients with pulvinar lesions differ from those of healthy individuals? If so, do patients with pulvinar lesions consistently display ignition events, or do these events show changes in frequency or characteristics?

What mechanisms does the pulvinar employ to adjust cortico-cortical connections when there is a mismatch between the predicted visual stimuli and the actual visual experiences of an individual? How does this adjustment process influence perception, and what are the implications when these mechanisms malfunction, particularly in relation to visual disturbances observed in disorders like schizophrenia?

Do transthalamic connections in biological neural networks share similarities with 'skip connections' observed in deep artificial neural networks? Skip connections feed the output of a certain layer to other non-adjacent layers within a network, bypassing intermediate layers. Skip connections are used to reduce signal dampening in deep neural networks and improve efficiency when the complexity of a hierarchical network increases. By using skip connections to model the pulvinar, could researchers gain insights into how this region functions and how it interacts with other regions of the brain? Moreover, given the pulvinar's complex role, is it feasible for it to act both as a central hub with

219
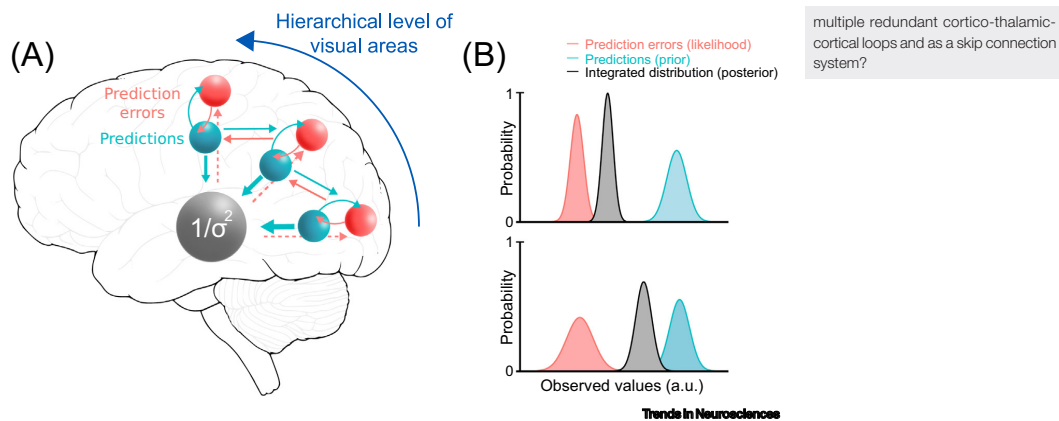
Figure 3. Predictive coding along the visual hierarchy. (A) Schematic illustration of the visual hierarchy under the predictive coding principle, in which locally distributed feedback predictions (depicted in blue) are matched against visual input to produce prediction errors (depicted in red), updating the internal model of the brain. This model is a generative one, allowing the brain to perform inference on the causes of the sensory inputs, by mapping external causes in the environment to generate internal resultant sensory experiences. Putatively, predictions are sent from the cortex to the pulvinar, where the inverse squared variance (i.e., the precision) of cortical signals can serve to modulate the propagation of messages based on their reliability, allowing information irrelevant to the brain's models to be disregarded. (B) Integration of prediction error and prediction into a posterior distribution, for high and low precision input (top and bottom row, respectively), under the approximation of Gaussian distribution of signals. This integration rule, postulated by Bayes' theorem (Box 2), drives the integrated posterior distribution towards (or away) the lower levels of the visual hierarchy, depending on the inverse variance of sensory inputs.

research avenue. Such studies may inform the design of more advanced, sophisticated, and efficient artificial intelligence algorithms inspired by transthalamic anatomical and functional mechanisms. Finally, specific involvement in disorders like schizophrenia, especially concerning 'ignition events' (those critical episodes of neural activation or synchronization), promises to yield valuable insights about the role of the pulvinar in such disorders and possibly target therapeutic interventions (see Outstanding questions). Addressing these questions will help decipher the role of the pulvinar in the complexity of brain functions and may offer promising insights that will deepen current understanding of cognitive processes, neural networks' functioning, and brain information processing.

Box 3. Does pulvinar malfunction contribute to mental disorders like schizophrenia?

Among the various neuropsychiatric disorders suspected to involve pulvinar malfunction, schizophrenia stands out as one of the most extensively studied and well-documented conditions [8,133,134]. Schizophrenia is characterized by disturbances in perceptual processing (e.g., visual hallucinations) and cognitive functions (e.g., impaired working memory, reduced processing speed), some of which align closely with the key processes in which the pulvinar normally participates. In a study analyzing postmortem thalamus samples from schizophrenia patients using unbiased stereological methods, structural changes such as reduced volume and cell numbers were consistently reported, particularly in the pulvinar's medial subdivision [135]. Additionally, functional connectivity studies using neuroimaging techniques in people with schizophrenia have revealed reduced connectivity between the PM and the frontal cortex [114,134]. Similar findings have been observed in early psychosis subjects and persist in individuals with chronic schizophrenia. These structural changes have been associated with impaired emotional processing [136] deficits in selective attention [137,138], abnormal motion processing [139], and impaired face recognition [140], functions known to involve the pulvinar. While these observations strongly suggest a link between pulvinar dysfunction and schizophrenia, it is important to note that schizophrenia is a complex disorder with multifactorial origins. The precise mechanisms by which pulvinar dysfunction contributes to the development and progression of schizophrenia remain a subject of ongoing research. Further investigations are needed to unravel the intricate interplay between pulvinar abnormalities and the complex pathophysiology of schizophrenia.

References

1. Goodale, M.A. (2010) Transforming vision into action. *Vis. Res.* 51, 1567–1587
2. Ungerleider, L.G. and Mishkin, M. (1982) Two cortical visual systems. In *Analysis of Visual Behavior* (Ingle, D.J. *et al.*, eds), pp. 549–586, The MIT Press
3. Casanova, C. (2021) Pulvinar and transthalamic cortical visual pathways. In *The Senses: A Comprehensive Reference* (2nd edn) (Fritzsch, B., ed.), pp. 472–479, Elsevier
4. Sherman, S.M. (2017) Functioning of circuits connecting thalamus and cortex. *Compr. Physiol.* 7, 713–739
5. Froesel, M. *et al.* (2021) A multisensory perspective onto primate pulvinar functions. *Neurosci. Biobehav. Rev.* 125, 231–243
6. Mease, R.A. and Gonzalez, A.J. (2021) Corticothalamic pathways from layer 5: emerging roles in computation and pathology. *Front. Neural Circuits* 15, 730211
7. Benarroch, E.E. (2015) Pulvinar: associative role in cortical function and clinical correlations. *Neurology* 84, 738–747
8. Homman-Ludiye, J. and Bourne, J.A. (2019) The medial pulvinar: function, origin and association with neurodevelopmental disorders. *J. Anat.* 235, 507–520
9. Shen, L. *et al.* (2021) Hypothesis of subcortical visual pathway impairment in schizophrenia. *Med. Hypotheses* 156, 110686
10. Chalfin, B.P. *et al.* (2007) Scaling of neuron number and volume of the pulvinar complex in New World primates: comparisons with humans, other primates, and mammals. *J. Comp. Neurol.* 504, 265–274
11. Arcaro, M.J. *et al.* (2015) The anatomical and functional organization of the human visual pulvinar. *J. Neurosci.* 35, 9848–9871
12. Baldwin, M.K.L. *et al.* (2017) The evolution and functions of nuclei of the visual pulvinar in primates. *J. Comp. Neurol.* 525, 3207–3226
13. Kaas, J.H. and Baldwin, M.K.L. (2019) The evolution of the pulvinar complex in primates and its role in the dorsal and ventral streams of cortical processing. *Vision (Basel)* 4, 3
14. Baldwin, M.K.L. and Bourne, J.A. (2020) The evolution of subcortical pathways to the extrastriate cortex. In *Evolutionary Neuroscience* (2nd edn) (Kaas, J.H., ed.), pp. 565–587, Academic Press
15. Kaas, J.H. and Lyon, D.C. (2007) Pulvinar contributions to the dorsal and ventral streams of visual processing in primates. *Brain Res. Rev.* 55, 285–296
16. Bridge, H. *et al.* (2016) Adaptive pulvinar circuitry supports visual cognition. *Trends Cogn. Sci.* 20, 146–157
17. Vittek, A.L. *et al.* (2023) Multisensory integration in neurons of the medial pulvinar of macaque monkey. *Cereb. Cortex* 33, 4202–4215
18. Abbas Farishta, R. *et al.* (2020) Hierarchical organization of corticothalamic projections to the pulvinar. *Cereb. Cortex Commun.* 1, tgaa030
19. Huppe-Gourgues, F. *et al.* (2006) Distribution, morphology, and synaptic targets of corticothalamic terminals in the cat lateral posterior-pulvinar complex that originate from the posteromedial lateral suprasylvian cortex. *J. Comp. Neurol.* 497, 847–863
20. Chalupa, L.M. and Abramson, B.P. (1989) Visual receptive fields in the striate-recipient zone of the lateral posterior-pulvinar complex. *J. Neurosci.* 9, 347–357
21. Kelly, L.R. *et al.* (2003) Ultrastructure and synaptic targets of tectothalamic terminals in the cat lateral posterior nucleus. *J. Comp. Neurol.* 464, 472–486
22. Boire, D. *et al.* (2004) Retinal projections to the lateral posterior-pulvinar complex in intact and early visual cortex lesioned cats. *Exp. Brain Res.* 159, 185–196
23. Zhou, N.A. *et al.* (2017) The mouse pulvinar nucleus: organization of the tectorecipient zones. *Vis. Neurosci.* 34, E011
24. Zhou, N. *et al.* (2018) The mouse pulvinar nucleus links the lateral extrastriate cortex, striatum, and amygdala. *J. Neurosci.* 38, 347–362
25. Arcaro, M.J. *et al.* (2018) Organizing principles of pulvino-cortical functional coupling in humans. *Nat. Commun.* 9, 5382
26. Kinoshita, M. *et al.* (2019) Dissecting the circuit for blindsight to reveal the critical role of pulvinar and superior colliculus. *Nat. Commun.* 10, 135
27. Casanova, C. *et al.* (1989) Monocular and binocular response properties of cells in the striate-recipient zone of the cat's lateral posterior-pulvinar complex. *J. Neurophysiol.* 62, 544–557
28. Bender, D.B. (1982) Receptive-field properties of neurons in the macaque inferior pulvinar. *J. Neurophysiol.* 48, 1–17
29. Dumbrava, D. *et al.* (2001) Global motion integration in the cat's lateral posterior-pulvinar complex. *Eur. J. Neurosci.* 13, 2218–2226
30. Merabet, L. *et al.* (1998) Motion integration in a thalamic visual nucleus. *Nature* 396, 265–268
31. Villeneuve, M.Y. *et al.* (2012) Pattern-motion selective responses in MT, MST and the pulvinar of humans. *Eur. J. Neurosci.* 36, 2849–2858
32. Moore, B. *et al.* (2019) Cortical projections to the two retinotopic maps of primate pulvinar are distinct. *J. Comp. Neurol.* 527, 577–588
33. Mundinano, I.C. *et al.* (2019) Retinotopic specializations of cortical and thalamic inputs to area MT. *Proc. Natl. Acad. Sci. U. S. A.* 116, 23326–23331
34. Hutchins, B. and Updyke, B.V. (1989) Retinotopic organization within the lateral posterior complex of the cat. *J. Comp. Neurol.* 285, 350–398
35. Schneider, L. *et al.* (2020) Eye position signals in the dorsal pulvinar during fixation and goal-directed saccades. *J. Neurophysiol.* 123, 367–391
36. Bertini, C. *et al.* (2018) Pulvinar lesions disrupt fear-related implicit visual processing in hemianopic patients. *Front. Psychol.* 9, 2329
37. Le, Q.V. *et al.* (2016) Snakes elicit earlier, and monkey faces, later, gamma oscillations in macaque pulvinar neurons. *Sci. Rep.* 6, 20595
38. Celeghin, A. *et al.* (2015) From affective blindsight to emotional consciousness. *Conscious. Cogn.* 36, 414–425
39. Kaas, J.H. (2015) Blindsight: post-natal potential of a transient pulvinar pathway. *Curr. Biol.* 25, R155–R157
40. de Souza, B.O.F. *et al.* (2021) Laminar distribution of cortical projection neurons to the pulvinar: a comparative study in cats and mice. *J. Comp. Neurol.* 529, 2055–2069
41. Van Le, Q. *et al.* (2013) Pulvinar neurons reveal neurobiological evidence of past selection for rapid detection of snakes. *Proc. Natl. Acad. Sci. U. S. A.* 110, 19000–19005
42. Sherman, S.M. and Guillery, R.W. (1998) On the actions that one nerve cell can have on another: distinguishing "drivers" from "modulators". *Proc. Natl. Acad. Sci. U. S. A.* 95, 7121–7126
43. Rockland, K.S. (1996) Two types of corticopulvinar terminations: round (type 2) and elongate (type 1). *J. Comp. Neurol.* 368, 57–87
44. Rockland, K.S. (2019) Corticothalamic axon morphologies and network architecture. *Eur. J. Neurosci.* 49, 969–977
45. Bickford, M.E. (2015) Thalamic circuit diversity: modulation of the driver/modulator framework. *Front. Neural Circuits* 9, 86

221

46. Kirchgessner, M.A. *et al.* (2021) Distinct "driving" versus "modulatory" influences of different visual corticothalamic pathways. *Curr. Biol.* 31, 5121–5137

47. Cortes, N. *et al.* (2021) Corticothalamic projections gate alpha rhythms in the pulvinar. *Front. Cell. Neurosci.* 15, 787170

48. Miura, S.K. and Scanziani, M. (2022) Distinguishing externally from saccade-induced motion in visual cortex. *Nature* 610, 135–142

49. Crick, F. and Koch, C. (1998) Constraints on cortical and thalamic projections: the no-strong-loops hypothesis. *Nature* 391, 245–250

50. Sherman, S.M. (2016) Thalamus plays a central role in ongoing cortical functioning. *Nat. Neurosci.* 19, 533–541

51. Bender, D.B. (1983) Visual activation of neurons in the primate pulvinar depends on cortex but not colliculus. *Brain Res.* 279, 258–261

52. Casanova, C. *et al.* (1997) Contribution of area 17 to cell responses in the striate-recipient zone of the cat's lateral posterior-pulvinar complex. *Eur. J. Neurosci.* 9, 1026–1036

53. de Souza, B.O.F. *et al.* (2020) Pulvinar modulates contrast responses in the visual cortex as a function of cortical hierarchy. *Cereb. Cortex* 30, 1068–1086

54. Purushothaman, G. *et al.* (2012) Gating and control of primary visual cortex by pulvinar. *Nat. Neurosci.* 15, 905–912

55. Clascá, F. *et al.* (2016) Anatomy and development of multispecific thalamocortical axons: implications for cortical dynamics and evolution. In *Axons and Brain Architecture* (Rockland, K.S., ed.), pp. 69–92, Academic Press

56. Rockland, K.S. (2021) A closer look at corticothalamic "loops". *Front. Neural Circuits* 15, 632668

57. Abbott, L.F. and Chance, F.S. (2005) Drivers and modulators from push-pull and balanced synaptic input. *Prog. Brain Res.* 149, 147–155

58. William, L.L. (2022) Comparison of additive and multiplicative feedforward control. *J. Process Control* 111, 1–7

59. Li, J. *et al.* (2003) Two distinct types of corticothalamic EPSPs and their contribution to short-term synaptic plasticity. *J. Neurophysiol.* 90, 3429–3440

60. Bastos, A.M. *et al.* (2015) Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* 85, 390–401

61. Jensen, O. *et al.* (2015) Oscillatory mechanisms of feedforward and feedback visual processing. *Trends Neurosci.* 38, 192–194

62. van Kerkoerle, T. *et al.* (2014) Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 111, 14332–14341

63. Vanni, M.P. *et al.* (2015) Spatiotemporal profile of voltage-sensitive dye responses in the visual cortex of tree shrews evoked by electric microstimulation of the dorsal lateral geniculate and pulvinar nuclei. *J. Neurosci.* 35, 11891–11896

64. Soares, J.G. *et al.* (2004) Effects of inactivation of the lateral pulvinar on response properties of second visual area cells in *Cebus* monkeys. *Clin. Exp. Pharmacol. Physiol.* 31, 580–590

65. Kohn, A. *et al.* (2020) Principles of corticocortical communication: proposed schemes and design considerations. *Trends Neurosci.* 43, 725–737

66. Shipp, S. (2003) The functional logic of cortico-pulvinar connections. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 358, 1605–1624

67. Worden, R. *et al.* (2021) The thalamus as a blackboard for perception and planning. *Front. Behav. Neurosci.* 15, 633872

68. Petersen, S.E. *et al.* (1985) Pulvinar nuclei of the behaving rhesus monkey: visual responses and their modulation. *J. Neurophysiol.* 54, 867–886

69. Petersen, S.E. *et al.* (1987) Contributions of the pulvinar to visual spatial attention. *Neuropsychologia* 25, 97–105

70. Arend, I. *et al.* (2008) Spatial and temporal deficits are regionally dissociable in patients with pulvinar lesions. *Brain* 131, 2140–2152

71. Fischer, J. and Whitney, D. (2012) Attention gates visual coding in the human pulvinar. *Nat. Commun.* 3, 1051

72. Gattass, R. *et al.* (2018) The role of the pulvinar in spatial visual attention. *Adv. Anat. Embryol. Cell Biol.* 225, 57–60

73. Kastner, S. *et al.* (2020) Dynamic pulvino-cortical interactions in the primate attention network. *Curr. Opin. Neurobiol.* 65, 10–19

74. Mizzi, R. and Michael, G.A. (2016) Exploring visual attention functions of the human extrageniculate pathways through behavioral cues. *Psychol. Rev.* 123, 740–757

75. Saalmann, Y.B. and Kastner, S. (2011) Cognitive and perceptual functions of the visual thalamus. *Neuron* 71, 209–223

76. Snow, J.C. *et al.* (2009) Impaired attentional selection following lesions to human pulvinar: evidence for homology between human and monkey. *Proc. Natl. Acad. Sci. U. S. A.* 106, 4054–4059

77. Strumpf, H. *et al.* (2013) The role of the pulvinar in distractor processing and visual search. *Hum. Brain Mapp.* 34, 1115–1132

78. Guedj, C. and Vuilleumier, P. (2023) Modulation of pulvinar connectivity with cortical areas in the control of selective visual attention. *NeuroImage* 266, 119832

79. Saalmann, Y.B. *et al.* (2012) The pulvinar regulates information transmission between cortical areas based on attention demands. *Science* 337, 753–756

80. Zhou, H. *et al.* (2016) Pulvinar-cortex interactions in vision and attention. *Neuron* 89, 209–220

81. Yu, C. *et al.* (2018) Theta oscillations organize spiking activity in higher-order visual thalamus during sustained attention. *eNeuro* 5, ENEURO.0384-17.2018

82. Cortes, N. and van Vreeswijk, C. (2012) The role of pulvinar in the transmission of information in the visual hierarchy. *Front. Comput. Neurosci.* 6, 29

83. Cortes, N. and van Vreeswijk, C. (2015) Pulvinar thalamic nucleus allows for asynchronous spike propagation through the cortex. *Front. Comput. Neurosci.* 9, 60

84. Jaramillo, J. *et al.* (2019) Engagement of pulvino-cortical feedforward and feedback pathways in cognitive computations. *Neuron* 101, 321–336

85. Miller-Hansen, A.J. and Sherman, S.M. (2022) Conserved patterns of functional organization between cortex and thalamus in mice. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2201481119

86. Cortes, N. *et al.* (2020) Pulvinar modulates synchrony across visual cortical areas. *Vision (Basel)* 4, 22

87. Eradath, M.K. *et al.* (2021) A causal role for the pulvinar in coordinating task-independent cortico-cortical interactions. *J. Comp. Neurol.* 529, 3772–3784

88. Fiebelkorn, I.C. *et al.* (2019) The mediodorsal pulvinar coordinates the macaque fronto-parietal network during rhythmic spatial attention. *Nat. Commun.* 10, 215

89. Halgren, M. *et al.* (2019) The generation and propagation of the human alpha rhythm. *Proc. Natl. Acad. Sci. U. S. A.* 116, 23772–23782

90. Hu, F. *et al.* (2019) Prefrontal corticotectal neurons enhance visual processing through the superior colliculus and pulvinar thalamus. *Neuron* 104, 1141–1152

91. Jones, E.G. (2002) Thalamic circuitry and thalamocortical synchrony. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 357, 1659–1673

92. Wilke, M. *et al.* (2009) Neural activity in the visual thalamus reflects perceptual suppression. *Proc. Natl. Acad. Sci. U. S. A.* 106, 9465–9470

93. Baars, B.J. (2005) Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Prog. Brain Res.* 150, 45–53

94. Baars, B.J. *et al.* (2021) Global workspace theory (GWT) and prefrontal cortex: recent developments. *Front. Psychol.* 12, 749868

95. Dehaene, S. and Changeux, J.P. (2011) Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227

96. Melloni, L. *et al.* (2021) Making the hard problem of consciousness easier. *Science* 372, 911–912

97. van Vugt, B. *et al.* (2018) The threshold for conscious report: signal loss and response bias in visual and frontal cortex. *Science* 360, 537–542

98. Fish, S.E. and Chalupa, L.M. (1979) Functional properties of pulvinar-lateral posterior neurons which receive input from the superior colliculus. *Exp. Brain Res.* 36, 245–257

99. Sato, T.K. *et al.* (2012) Traveling waves in visual cortex. *Neuron* 75, 218–229

100. Logothetis, N.K. *et al.* (2010) The effects of electrical microstimulation on cortical signal propagation. *Nat. Neurosci.* 13, 1283–1291

101. Rockland, K.S. (2019) Distinctive spatial and laminar organization of single axons from lateral pulvinar in the macaque. *Vision (Basel)* 4, 1

222

102. Rockland, K.S. *et al.* (1999) Single axon analysis of pulvinocortical connections to several visual areas in the macaque. *J. Comp. Neurol.* 406, 221–250

103. Ramcharan, E.J. *et al.* (2005) Higher-order thalamic relays burst more than first-order relays. *Proc. Natl. Acad. Sci. U. S. A.* 102, 12236–12241

104. Blot, A. *et al.* (2021) Visual intracortical and transthalamic pathways carry distinct information to cortical areas. *Neuron* 109, 1996–2008

105. Kagan, I. *et al.* (2021) Effective connectivity and spatial selectivity-dependent fMRI changes elicited by microstimulation of pulvinar and LIP. *NeuroImage* 240, 118283

106. Klink, P.C. *et al.* (2017) Distinct feedforward and feedback effects of microstimulation in visual cortex reveal neural mechanisms of texture segregation. *Neuron* 95, 209–220

107. Chalupa, L.M. (1991) Visual function of the pulvinar. In *The Neural Basis of Visual Function* (Leventhal, A.G., ed.), pp. 140–159, Macmillan

108. Desautels, A. and Casanova, C. (2001) Response properties in the pulvinar complex after neonatal ablation of the primary visual cortex. *Prog. Brain Res.* 134, 83–95

109. Rao, R.P. and Ballard, D.H. (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87

110. Mumford, D. (1991) On the computational architecture of the neocortex. I. The role of the thalamocortical loop. *Biol. Cybern.* 65, 135–145

111. Kanai, R. *et al.* (2015) Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 370, 20140169

112. Haarsma, J. *et al.* (2021) Precision weighting of cortical unsigned prediction error signals benefits learning, is mediated by dopamine, and is impaired in psychosis. *Mol. Psychiatry* 26, 5320–5333

113. Casanova, C. and Chalupa, L.M. (2023) The dorsal lateral geniculate nucleus and the pulvinar as essential partners for visual cortical functions. *Front. Neurosci.* 17, 1258393

114. Penner, J. *et al.* (2018) Higher order thalamic nuclei resting network connectivity in early schizophrenia and major depressive disorder. *Psychiatry Res. Neuroimaging* 272, 7–16

115. Beltramo, R. and Scanziani, M. (2019) A collicular visual cortex: neocortical space for an ancient midbrain visual structure. *Science* 363, 64–69

116. Grunert, U. *et al.* (2021) Retinal ganglion cells projecting to superior colliculus and pulvinar in marmoset. *Brain Struct. Funct.* 226, 2745–2762

117. Leow, Y.N. *et al.* (2022) Brain-wide mapping of inputs to the mouse lateral posterior (LP/Pulvinar) thalamus-anterior cingulate cortex network. *J. Comp. Neurol.* 530, 1992–2013

118. Kwan, W.C. *et al.* (2019) Unravelling the subcortical and retinal circuitry of the primate inferior pulvinar. *J. Comp. Neurol.* 527, 558–576

119. Ito, S. and Feldheim, D.A. (2018) The mouse superior colliculus: an emerging model for studying circuit formation and function. *Front. Neural Circuits* 12, 10

120. Wassle, H. and Illing, R.B. (1980) The retinal projection to the superior colliculus in the cat: a quantitative study with HRP. *J. Comp. Neurol.* 190, 333–356

121. Dhande, O.S. and Huberman, A.D. (2014) Retinal ganglion cell maps in the brain: implications for visual processing. *Curr. Opin. Neurobiol.* 24, 133–142

122. Brenner, J.M. *et al.* (2023) A genetically defined tecto-thalamic pathway drives a system of superior-colliculus-dependent visual cortices. *Neuron* 111, 2247–2257

123. Balaram, P. *et al.* (2015) Distributions of vesicular glutamate transporters 1 and 2 in the visual system of tree shrews (*Tupaia belangeri*). *J. Comp. Neurol.* 523, 1792–1808

124. Abbas Farishta, R. *et al.* (2022) Distributions of vesicular glutamate transporters 1 and 2 in the visual thalamus and associated areas of the cat. *J. Comp. Neurol.* 530, 1112–1125

125. Allen, A.E. *et al.* (2016) Visual input to the mouse lateral posterior and posterior thalamic nuclei: photoreceptive origins and retinotopic order. *J. Physiol.* 594, 1911–1929

126. Abramson, B.P. and Chalupa, L.M. (1985) The laminar distribution of cortical connections with the tecto- and cortico-recipient zones in the cat's lateral posterior nucleus. *Neuroscience* 15, 81–95

127. Abramson, B.P. and Chalupa, L.M. (1988) Multiple pathways from the superior colliculus to the extrageniculate visual thalamus of the cat. *J. Comp. Neurol.* 271, 397–418

128. Friston, K. (2005) A theory of cortical responses. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 360, 815–836

129. Aitchison, L. and Lengyel, M. (2017) With or without you: predictive coding and Bayesian inference in the brain. *Curr. Opin. Neurobiol.* 46, 219–227

130. Bastos, A.M. *et al.* (2012) Canonical microcircuits for predictive coding. *Neuron* 76, 695–711

131. Friston, K.J. (2019) Waves of prediction. *PLoS Biol.* 17, e3000426

132. Pezzulo, G. *et al.* (2022) The evolution of brain architectures for predictive coding and active inference. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 377, 20200531

133. Dorph-Petersen, K.A. and Lewis, D.A. (2017) Postmortem structural studies of the thalamus in schizophrenia. *Schizophr. Res.* 180, 28–35

134. Yamamoto, M. *et al.* (2018) Aberrant functional connectivity between the thalamus and visual cortex is related to attentional impairment in schizophrenia. *Psychiatry Res. Neuroimaging* 278, 35–41

135. Highley, J.R. *et al.* (2003) Low medial and lateral right pulvinar volumes in schizophrenia: a postmortem study. *Am. J. Psychiatry* 160, 1177–1179

136. Butler, P.D. *et al.* (2009) Sensory contributions to impaired emotion processing in schizophrenia. *Schizophr. Bull.* 35, 1095–1107

137. Mathalon, D.H. *et al.* (2004) Selective attention in schizophrenia: sparing and loss of executive control. *Am. J. Psychiatry* 161, 872–881

138. John, Y.J. *et al.* (2018) Visual attention deficits in schizophrenia can arise from inhibitory dysfunction in thalamus or cortex. *Comput. Psychiatr.* 2, 223–257

139. Martinez, A. *et al.* (2018) Impaired motion processing in schizophrenia and the attenuated psychosis syndrome: etiological and clinical implications. *Am. J. Psychiatry* 175, 1243–1254

140. Sachs, G. *et al.* (2004) Facial recognition deficits and cognition in schizophrenia. *Schizophr. Res.* 68, 27–35

223

C. Appendix C: Public Disseminations

"T'as pas de doutes ? Moi j'en ai,
(...)
Je sais pas ce que c'est,
C'est l'Inconnu."
Cadillac & KingJu, Egoslave, 2018

C.1. Rationale and open-access

There are several reasons why public research should be made publicly available. On the professional side, having any scientist able to access in-depth research, and not just the research manuscript, vastly improves the quality of any scientific field as a whole. On the ethical side, having an open policy on research effectively deprives predatory publishing companies of their revenues, vastly improving the quality of life of all scientists [290]. For these two reasons, when possible (i.e. not under embargo for a publication in production), the code that has been written during this thesis and its associated data is available online for open access:

— Related to this entire manuscript:
  Code: https://github.com/hugoladret/PhD_manuscript
— Related to Chapter's 3 article on Convolutional Sparse Coding:
  Data: https://doi.org/10.6084/m9.figshare.24167265
  Code: https://github.com/hugoladret/epistemic_CSC
— Related to Chapter's 4 article on Cortical Recurrence in V1:
  Data: https://doi.org/10.6084/m9.figshare.23366588.v2
  Code: https://github.com/hugoladret/variance-processing-V1

Finally, on the logical side, since this research is paid by taxpayer's money, it should be made available to the taxpayer. Taxes are not just paid by researchers (thankfully for us), but by people with a heterogeneous scientific formation, and as such, it is crucial that scientific productions end up being formatted in such a way that anyone might benefit from them. In that regard, this thesis also includes two public dissemination articles in French, with English translation provided here after each article.

C.2. Article 1 (Sciences et Avenir)

The first public dissemination article is based on an interview by Alice Carliez, derived from our article Hugo J Ladret, Nelson Cortes, Lamyae Ikan, et al. "Cortical recurrence supports resilience to sensory variance in the primary visual cortex". In: *Communications Biology* 6.1 (2023), p. 667, for the French journal *"Sciences et Avenir"*.

Comment notre cerveau fait-il face à l'incertitude ?

Par <u>Alice Carliez</u> le <u>28.07.2023 à 12h29</u>
Lecture 8 min.

Une équipe du CNRS et d'Aix-Marseille Université a élucidé les mécanismes neuronaux permettant la perception de stimuli visuels plus ou moins précis. Voici les explications de Laurent Perrinet, chercheur en neurosciences computationnelles.
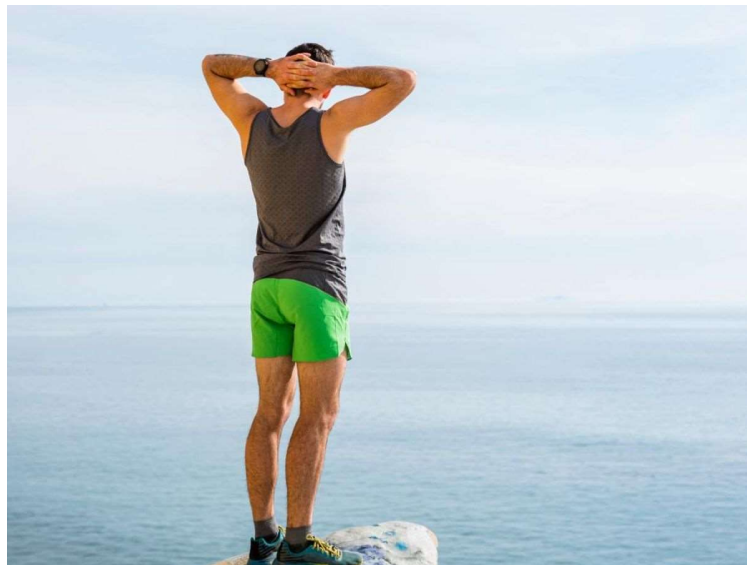
Comment le cerveau fait face à l'incertitude ?
©JFCREATIVES / CULTURA CREATIVE / CULTURA CREATIVE VIA AFP

"L'incertitude n'est pas dans les choses mais dans notre tête". Une récente étude pourrait concrètement confirmer cette assertion du mathématicien suisse Jacques Bernoulli.

Des images de synthèse pour représenter le degré de flou

Nous sommes au quotidien confronté à des environnement visuels complexes et à un ensemble de stimuli sensoriels à intégrer avant de prendre une décision. Le traitement de multiples informations en parallèle permet à notre cerveau d'adapter notre comportement à des situations très diverses. Cependant, le fonctionnement neurobiologique sous-jacent est encore mal compris. Dans une <u>étude publiée le 23 juin 2023 dans *Nature Communications Biology*</u>, des scientifiques ont utilisé des technologies novatrices pour comprendre l'activité des neurones du cortex visuel primaire en réponse à la présentation d'images reproduisant des situations d'incertitude. *Sciences et Avenir* s'est entretenu avec Laurent Perrinet, chercheur en neurosciences computationnelles dans l'équipe de <u>l'Institut de Neurosciences de la Timone </u>(CNRS / Aix-Marseille Université) et encadrant de thèse de Hugo Ladret, premier auteur de l'étude.

Reproduire dans une expérience de neurosciences la complexité du monde dans lequel on vit, tout en pouvant analyser les résultats de manière fiable, est un véritable défi. Laurent Perrinet a proposé d'utiliser des images de synthèse conçues pour reproduire

225

un contexte visuel incertain. A l'instar des textures utilisées dans les jeux vidéo en deux dimensions, ces images représentent des motifs allongés, orientés plus ou moins dans la même direction. Quand tous les motifs présentent la même orientation, il est très facile de la deviner. En revanche, lorsque de nombreuses orientations sont mélangées, l'information visuelle est beaucoup moins claire.

Ces textures permettent de reproduire les images naturelles auxquelles nous sommes confrontés. Tandis que nous arrivons parfois à très bien identifier un objet, d'autres éléments perçus nous paraissent plus flous, plus incertains. Dans l'utilisation des textures, le problème est le même : il y a tantôt des lignes dont l'orientation est facilement identifiable, et parfois il y a des points dispersés dont on peine à voir une organisation précise.

L'utilisation de ces textures est novatrice. En effet, classiquement, les chercheurs en neurosciences utilisaient plutôt des formes assez isolées dans la stimulation des aires visuelles : un rectangle, un point, une ligne qui se déplace. *"Le fait d'avoir des images aussi simples, c'est pratique pour faire des analyses. Mais ces formes isolées ne sont pas 'écologiques'. Il faut alors trouver une autre façon de procéder pour reproduire notre monde plus complexe. D'autant plus que le cerveau est adapté à percevoir des images naturelles très riches",* explique à *Sciences et Avenir* Laurent Perrinet.

Le cerveau est fait pour percevoir à la fois des objets précis : leur forme, direction, orientation, contours, couleurs... Mais aussi pour comprendre des situations plus incertaines, afin d'interpréter le désordre et le chaos qui troublent nos anticipations. Les textures obtenues par images de synthèse sont une manière de reproduire les images naturelles auxquelles nous sommes confrontés. En partant d'une intuition théorique mathématique, l'équipe de Laurent Perrinet s'est tout d'abord demandé comment le cerveau peut laisser une [place à l'incertitude](.).
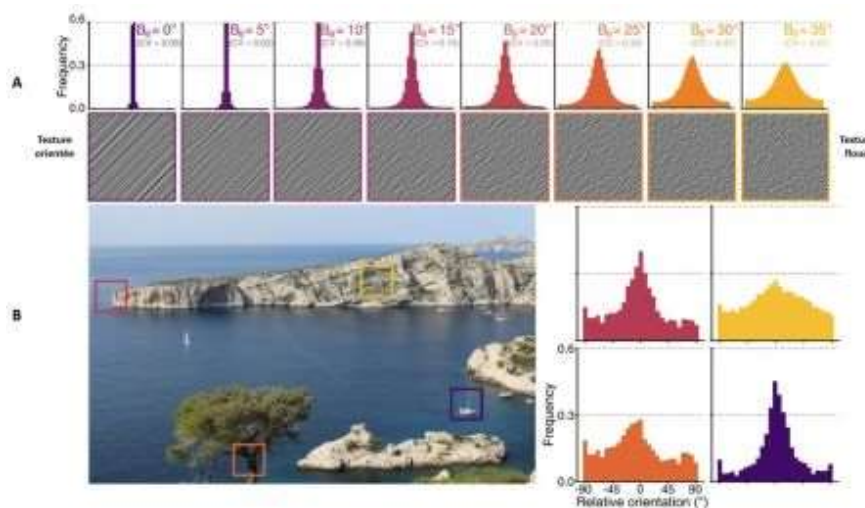
A. *Exemple de textures présentées lors des expériences. Le caractère incertain des images est associé au nombre d'orientations mélangées dans une images exprimée en degré, que l'on peut représenter sur un diagramme. Plus la courbe du diagramme est resserrée, plus*

les orientations sont similaires. Plus le diagramme est étalé horizontalement, plus il y a de motifs différents mélangés sur l'image. A gauche, les textures dont l'orientation est la plus facilement distinguable. A droite, les images les plus incertaines.

__B__. Les textures peuvent représenter les images naturelles auxquelles nous sommes confrontés au quotidien. Par exemple, si on réalise un diagramme représentant les distributions de l'orientation de quatre régions sur une image naturelle (photo prise par Hugo Ladret), on observe qu'il y a des régions dont l'orientation des motifs est claire, et des régions plus floues.

Crédit illustration : Hugo J. Ladret / Nature Communications Biology (2023) / l'Institut de Neurosciences de la Timone (CNRS et Aix-Marseille Université)

Des neurones spécialisés dans l'interprétation de l'imprécision

Les scientifiques ont enregistré l'activité de 249 neurones de chats anesthésiés. Ils ont observé la réponse des neurones dans l'aire visuelle primaire à la présentation d'images plus ou moins troubles. Ils ont observé qu'il y avait deux types de réponses neuronales à l'incertitude.

Les neurones "vulnérables" qui ne répondent que pour une certaine orientation. Ils sont très sensibles et vulnérables aux grands degrés de flou. Et les neurones "résistants" qui répondent aux stimuli visuels malgré le manque de précisions de l'information visuelle. Même lorsqu'on présente aux animaux des textures qui ne sont pas bien définies, c'est-à-dire dont l'orientation n'est pas distinguable, ces neurones continuent de répondre.

"Si on montre à une personne des textures qui figurent parmi les plus imprécises au sein de la gamme de texture, à partir de 30° d'imprécision, les personnes se trompent et ne trouvent pas l'orientation des lignes de la texture. Mais il y a tout de même des neurones qui répondent précisément !", détaille Laurent Perrinet.

"Je te parie que dans le cerveau, il y a une représentation du flou"

Les chercheurs ont donc voulu aller plus loin que l'observation de l'activité des neurones. Ils ont regardé si, à partir de cette activité, il était possible de reconstruire le type de texture présenté en premier lieu. Ils ont réalisé ce décryptage grâce à des processus d'apprentissage machine similaires à ceux utilisés dans le [Deep Learning](). *"Tout est parti de discussions entre nous et d'un pari entre les scientifiques ''je te parie que dans le cerveau, il y a une représentation du flou'",* raconte le chercheur. L'expérience est menée de sorte à comprendre l'encodage de l'information visuelle, à partir de l'activité des neurones, grâce à une étape de "décodage". Ainsi, on peut définir trois étapes :

- L'encodage : l'information lumineuse causée par l'image est captée par les yeux.
- Le codage : l'enregistrement de l'activité des neurones qui perçoivent cette information.
- Le décodage : la traduction de l'activité neuronale via un programme de [Machine Learning]() pour retrouver l'orientation des lignes sur l'image présentée à la première étape.

A la bonne surprise des scientifiques, cette étape de décodage fonctionne particulièrement bien : le logiciel décodeur peut retrouver l'orientation des lignes de la texture présentée en premier lieu de manière robuste et très proche de la réalité. Par

ailleurs, lorsque l'image présentée était très brouillée, ils ont observé que le décodage est un peu moins exact et apparaît avec un court délai. Mais, malgré le flou, le décodage fonctionne tout de même étonnement bien.

Les résultats confirment que la population de neurones de l'aire visuelle primaire peut non seulement retrouver l'orientation d'un objet, mais également interpréter si l'information est précise ou non. Le cerveau est capable de se représenter les stimuli visuels imprécis et de distinguer à quel point une information visuelle est certaine ou pas. *"Pour un réseau de neurones, il est important de pouvoir décrypter à la fois la nature d'une information mais également sa précision, cela pour participer à la prise de décision. Le cerveau fonctionne sans relâche, avec des neurones qui travaillent en groupe pour échanger et intégrer un ensemble d'informations, ce qui peut prendre du temps avant de prendre une décision de manière consensuelle. D'où l'importance de l'incertitude pour donner plus ou moins de poids à certaines informations, pour faciliter la prise de décision",* explique Laurent Perrinet.

Le cerveau, une machine à prédire

"Le cerveau ne devrait pas être considéré comme un ordinateur mais comme une machine à prédire. Un ensemble de cellules qui veulent notre bien et notre survie, en prenant des décisions avec un modèle probabiliste", illustre Laurent Perrinet. Selon lui, il serait erroné de faire l'analogie entre le fonctionnement cérébral et celui d'un ordinateur qui fonctionne de manière séquentielle. La théorie du "cerveau prédictif" propose que les cellules neuronales fonctionnent en continu et en groupes, et ce, pour intégrer la multiplicité des informations sensorielles perçues, afin de prendre une décision. Pour que cette décision soit prise de la manière la plus fluide possible, malgré l'imprécision liée à nos sens, il serait nécessaire que les neurones fassent de la prédiction.

"On voit actuellement une révolution technologique dans le Machine Learning sur l'utilisation des réseaux profonds, ChatGPT etc... Ces technologies sont basées sur des réseaux de neurones qui permettent d'obtenir des performances extraordinaires, mais qui n'égaleront jamais celles du cerveau. Ce dernier consomme 5 à 20 Watts. Un GPU (Graphical Processing Unit, architecture alternative aux processeurs communément utilisés dans le Deep Learning, ndlr) *actuel c'est 600 Watts, et celui qui a* [battu le champion du monde de Go](#) *c'est 20 Mégawatts !,* poursuit le chercheur. *C'est merveilleux qu'on puisse aujourd'hui utiliser ces intelligences de deep learning en santé ou pour plein d'autres activités. Mais, je pense qu'il faut garder en tête que ces technologies restent aujourd'hui très sensibles à des attaques. Cela ne semble pas être un problème si une technologie a pour simple tâche de distinguer un chat d'un chien. En revanche, cela peut être dangereux si on utilise des* [IA en imagerie médicale](#) *et qu'une variabilité dans les données reçues puisse influer sur la véracité du diagnostic. Si on voulait que les* [IA aient l'efficacité du cerveau](#)*, il faudrait inclure dans chaque nœud de ce réseau, en plus des valeurs, leurs précisions. Au lieu d'utiliser les réseaux actuels qui vont fonctionner de manière analogique, on pourrait mettre en place un fonctionnement probabiliste".* Le chercheur souhaite donner une place à l'incertitude. Exactement comme le fait notre cerveau.

English translation:
How does our brain deal with uncertainty?
A team from CNRS and Aix-Marseille University has elucidated the neural mechanisms that enable the perception of more or less precise visual stimuli. Here are the explanations by Laurent Perrinet, a researcher in computational neuroscience.

Every day, we are confronted with complex visual environments and a range of sensory stimuli to integrate before making a decision. Processing multiple pieces of information in parallel enables our brain to adapt to a wide variety of situations. However, the underlying process is still poorly understood. In a study published in June 23 2023 in Nature Communications Biology, scientists have used innovative technologies to understand the activity of neurons in the primary visual cortex cortex in response to the presentation of images reproducing situations of uncertainty. Sciences et Avenir spoke to Laurent Perrinet, a computational neuroscientist computational neuroscience researcher in the team at the Institut de Neurosciences de la Timone (CNRS / Aix-Marseille University) and Hugo Ladret's thesis supervisor, first author of the study.

Reproducing the complexity of the world in which we live in a neuroscience experiment, while still being able to analyze the results reliably, is a real challenge. Laurent Perrinet has proposed the use of computer-generated images designed to reproduce an uncertain visual context. Like the textures used in two-dimensional video games, these images represent elongated patterns, oriented more or less in the same direction. When all the patterns have the same orientation, it's easy to guess which is their orientation. On the other hand, when many different orientations are present, visual information is much less clear. These textures allow us to reproduce the natural images we are confronted with. While we can sometimes identify an object very well, other perceived elements appear more blurred and uncertain. When it comes to using textures, the problem is the same: sometimes there are lines whose orientation is easily identifiable, and sometimes there are scattered dots whose organization is imprecise.

The use of these textures is innovative. Traditionally, neuroscience researchers have used rather isolated shapes to stimulate visual areas: a rectangle, a dot, a moving line. "The fact of having such simple images is useful for analysis. But these isolated shapes are not 'ecological'. So we have to find another way of reproducing our more complex world. Especially as the brain is adapted to perceiving very rich images," explains Laurent Perrinet to Sciences et Avenir.

The brain is designed to perceive both precise objects: their shape, direction, orientation, contours, colors... But also to understand more uncertain situations, in order to interpret the disorder and chaos that confuse our anticipations. Computer-generated textures are a way of reproducing the natural images we are confronted with. Starting from a mathematical theoretical intuition, Laurent Perrinet's team first asked themselves how the brain can leave room for uncertainty.

Some neurons are specialized in uncertainty Scientists recorded the activity of 249 neurons in anesthetized cats. They observed the response of neurons in the primary visual area to the presentation of more or less blurred images. They found that there

were two types of neuronal response to uncertainty. Vulnerable" neurons respond only to a certain orientation. They are highly sensitive and vulnerable to large degrees of vagueness. And "resistant" neurons, which respond to visual stimuli despite the lack of precision of the visual information. Even when animals are presented with textures that are not well-defined, i.e. whose orientation is indistinguishable, these neurons continue to respond. "If a person is shown textures that are among the most imprecise within the texture range, from 30° of imprecision, people get confused and can't find the orientation of the texture lines. But there are still neurons that respond precisely", explains Laurent Perrinet.

"I bet you that in the brain, there is a representation of the blur" The researchers therefore wanted to go further than simply observing neuron activity. They looked to see if, from this activity, it was possible to reconstruct the type of texture presented in the first place. They achieved this deciphering through machine learning processes similar to those used in Deep Learning. "It all started from discussions between us and a bet between the scientists "I bet you that in the brain, there is a representation of the blur'", says the researcher.The experiment is conducted in such a way as to understand the encoding of visual information, based on the activity of neurons, thanks to a "decoding" stage.Three stages can be defined:

— Encoding: the luminous information caused by the image is captured by the eyes.
— Coding: the recording by the neuronal activity that perceive this encoding.
— Decoding: the translation of the neuronal activity via a Machine Learning algorithm, to retrieve the orientation of the lines on the images presented to the eyes.

To the surprise of the scientists, this decoding step works extremely well: the decoding software developed here can retrieve the orientation of the lines of the texture with great performance. Furthermore, when the image was blurry, they observed that the decoding was also more blurry, and appeared with a delay. The results confirm that the population of the neurons in the primary visual cortex can not only retrieve the orientation of an object, but also interpret whether that information is precise or not. The brain is capable of representing imprecise visual stimuli, and distinguish whether one information is precise or not. "For a neural network, it is important to decipher both the nature and the precision of visual information, to be able to make a decision. The brain works tirelessly, with neurons working together to exchange and integrate packets of information, which can take time before reaching a consensus. Hence, the importance of uncertainty to weigh more or less some information, to save time and facilitate decision-making", explains Laurent Perrinet.

The brain, a predictive machine. "The brain should not be considered as a computer, but as a prediction machine. A set of cells that want what's best for us and work for our survival, making decisions that can be understood using a probabilistic model," illustrates Laurent Perrinet. In his view, it would be a mistake to draw an analogy between brain function and that of a sequentially operating computer. The theory of the "predictive brain" proposes that neuronal cells function continuously and in groups, to integrate the multiplicity of sensory information perceived, in order

to make a decision. In order for this decision to be taken as smoothly as possible, despite the imprecision associated with our senses, it would be necessary for neurons to make predictions.

"We're currently seeing a technological revolution in Machine Learning on the use of deep networks, ChatGPT etc.... These technologies are based on neural networks which enable extraordinary performance, but which are not yet equal to that of the brain. The latter consumes 5 to 20 Watts. A current GPU (Graphical Processing Unit, an alternative architecture to the processors commonly used in Deep Learning, editor's note) is 600 Watts, and the one that beat the Go world champion is 20 Megawatts! continues the researcher. It's marvellous that we can now use these deep learning intelligences in healthcare or for so many other activities. But I think we have to bear in mind that these technologies are still very sensitive to errors. This doesn't seem to be a problem if a technology's task is simply to distinguish a cat from a dog. On the other hand, it could be dangerous if AIs are used in medical imaging, where variability in the data received could influence the veracity of the diagnosis. If we want AIs to have the efficiency of the brain, we would have to include in each node of this network, in addition to the values, their uncertainty. Instead of using current networks that operate in an analogical way, we could implement a probabilistic operation". The researcher wants to give uncertainty a place. Just as our brains do.

C.3. Article 2 (Cerveau et Psycho)

The second public dissemination article was written by Laurent Perrinet and myself, based on our article Hugo J Ladret, Nelson Cortes, Lamyae Ikan, et al. "Cortical recurrence supports resilience to sensory variance in the primary visual cortex". In: *Communications Biology* 6.1 (2023), p. 667, to be published in the French journal *"Cerveau et Psycho".* Compared to the first dissemination article, this one dives (slightly) more into the philosophical implications of processing uncertainty, as well as providing more common sense into the notion of uncertainty in natural images.

Cerveau qui devine, cerveau qui doute
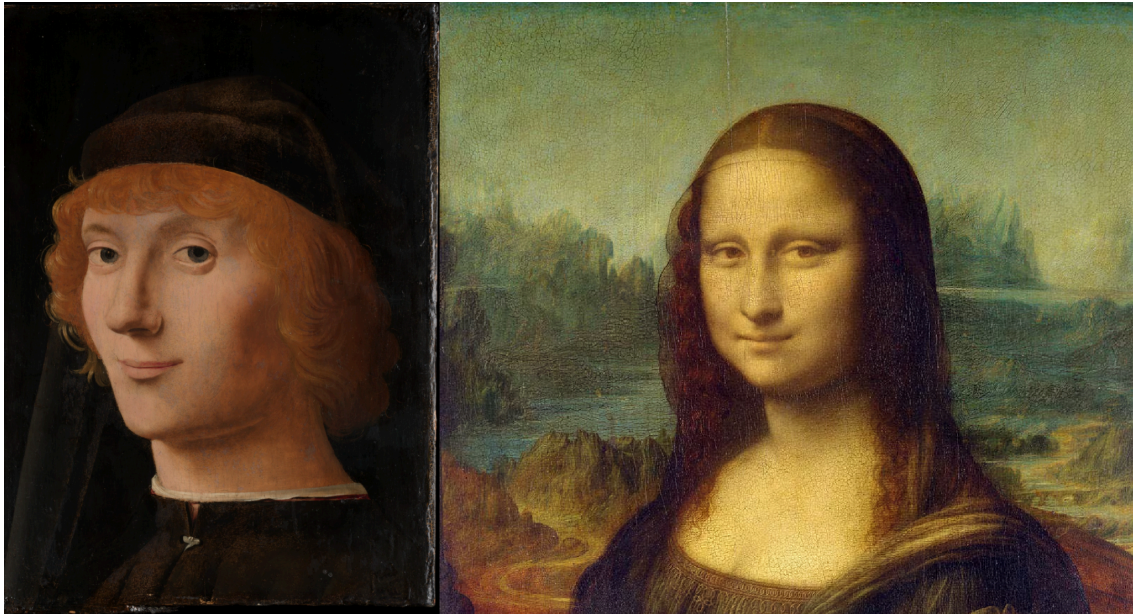
À gauche : Antonello da Messina, Portrait of a Young Man, 1470. À droite : Leonardo Da Vinci, Monna Lisa, 1503 - 1517

INTRO

Si l'on vous demande lequel de ces deux portraits donne le rendu le plus fidèle de l'expression d'un visage, vous répondrez très probablement qu'il s'agit de l'iconique Monna Lisa. Toutefois, si on vous demande maintenant lequel de ces deux visages dévoile de manière la plus claire et précise son émotion, vous trouverez sans doute plus aisé de lire les pensées rieuses rendues dans le portrait d'*Antonello da Messina*. Il semble donc que même si vous percevez Monna Lisa plus "vivante", elle ne vous en paraît pas plus compréhensible pour autant. Intrigant, n'est-ce pas ? Bien qu'il émane de ces deux personnages un même air de mystère, quelque chose confère à la Joconde une singularité envoutante.

À la lumière d'une analyse scientifique moderne, nous comprenons désormais mieux pourquoi le portrait de Monna Lisa paraît si authentique et si mystérieux. L'ambivalence légendaire de l'œuvre permet au regard de la fameuse Joconde de suivre le spectateur ainsi qu'à son sourire de paraître changer selon l'angle de vue. Da Vinci, fin observateur du monde qui l'entourait, a magistralement utilisé des techniques réalistes pour donner vie à une image qui, tout en étant statique, évoque une présence dynamique comparable à celles que nous rencontrons dans nos vies quotidiennes. L'ambiguïté visuelle induite par l'observation de la Joconde, un aspect central de son mystère, rappelle la manière dont la lumière du soleil se fraye un chemin à travers un feuillage dans une forêt, jouant subtilement avec les ombres pour créer une atmosphère à la fois vivante et énigmatique.

Cet énigmatique réalisme du portrait de Monna Lisa repose sur le talent de son créateur. Da Vinci était un maître incontesté du *sfumato*, une technique artistique ardue qui permet d'adoucir les contours dans une peinture, pour un effet plus naturel et réaliste. Il s'agit là d'un bel exemple de son génie pluridisciplinaire, car da Vinci, avec seulement les connaissances de son époque, semblait anticiper certaines avancées de la science moderne sur les images naturelles. En effet, nous savons désormais que notre cerveau est particulièrement efficace pour analyser sous toutes leurs coutures le monde visuel qui

nous entoure. Un système d'aires cérébrales traite notamment les informations en provenance de la rétine, en commençant par décomposer le monde qui nous entoure en petits éléments de contours, orientés et subtils, minutieusement reproduits dans le *sfumato* de la Joconde. Ces contours échafaudent en quelque sorte la charpente de notre perception visuelle, à partir de laquelle notre cerveau esquisse une image du monde lumineux qui nous entoure.

ANALYSE DES IMAGES NATURELLES
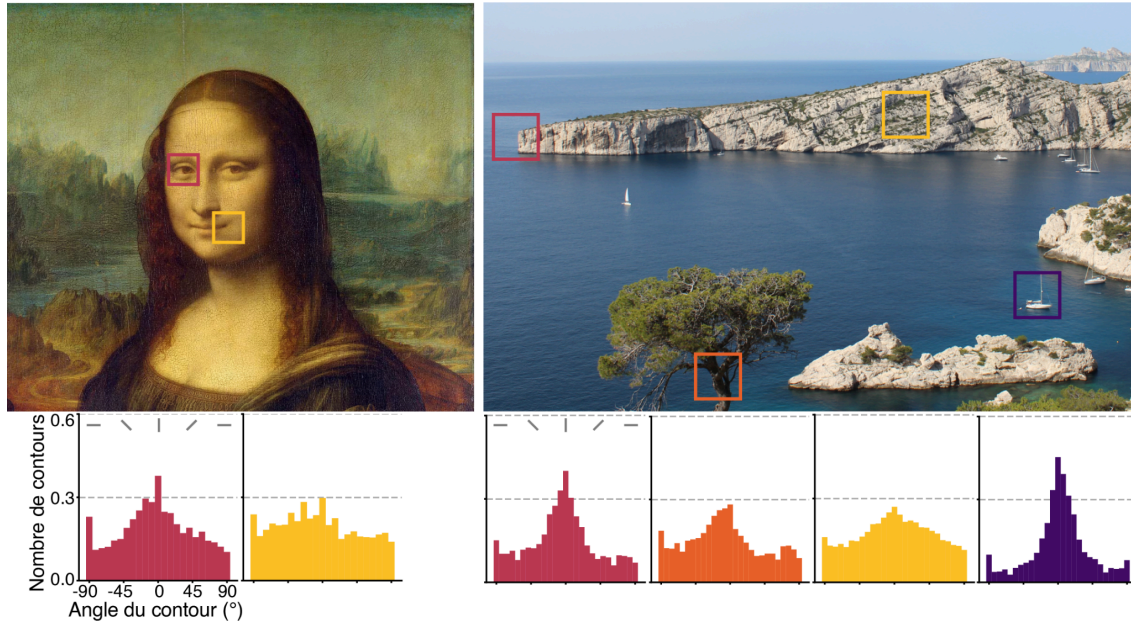
À gauche : décomposition de l'œil et du sourire de Monna Lisa en une série de contours orientés. À droite : même opération, pour une photographie des Calanques de Marseille.

Tout comme le sourire énigmatique de Monna Lisa, les images naturelles sont elles-mêmes imprégnées d'incertitude et de complexité. Prenons, par exemple, la différence flagrante entre les décompositions des orientations qui forment les contours d'un arbre et ceux qui dessinent un bateau telle qu'elle est représentée dans la figure ci-dessus. Les incertitudes différentes de ces contours mesurent des degrés d'ambiguïté fondamentalement distincts au sein d'une même image et illustrent parfaitement la variété inhérente à notre environnement visuel.

Dans un contexte moderne, ce constat prend toute son importance. Imaginez-vous être un piéton, en pleine ville, et vous apprêtant à traverser la rue à un passage sans feu. Vous fixez du regard le conducteur de la voiture la plus proche, et essayez de lire sur son visage ses intentions. Faut-il faire confiance à ses yeux qui vous ont fixés ? Ou bien prendre en compte l'incertitude de sa voiture qui a ralenti sans tout à fait freiner ? Faut-il attendre de la résolution de cette incertitude par un geste de sa main ? Autant de variables visuelles à considérer pour réduire l'incertitude quant au fait que le conducteur va s'arrêter, tout en espérant que l'air de ce dernier soit moins mystérieux que la Joconde.

CERVEAU QUI DEVINE

Il ne fait donc aucun doute que l'incertitude est une question fondamentale en "*psycho*". Mais qu'en est-il du "*cerveau*" lui-même ? Une étude récente que nous avons menée à l'Institut des Neurosciences de la Timone (Aix-Marseille Université; CNRS) en collaboration avec l'Université de Montréal, apporte des éclaircissements sur ce sujet.
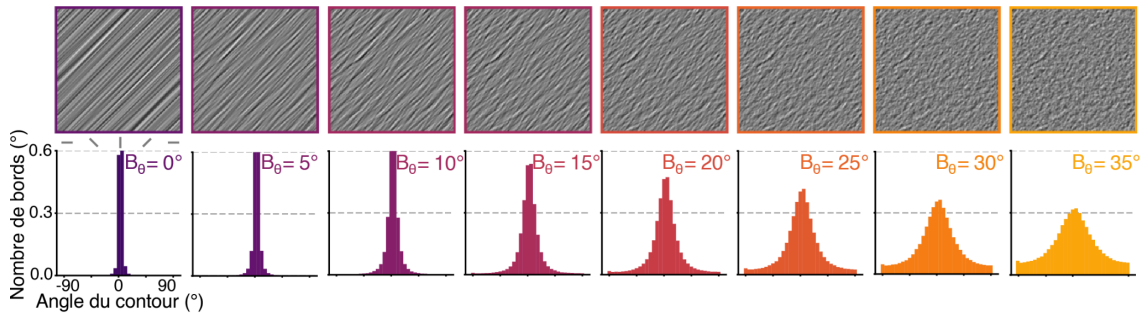

Dans notre recherche, nous avons utilisé une technique de génération d'images qui nous permet de créer des textures contrôlant différents degrés d'incertitude, allant d'une certitude absolue (les lignes parallèles à gauche) à l'ambiguïté du sfumato (la texture à droite).

En observant des neurones impliqués dans la vision, et en particulier dans la représentation de contours orientés, nous avons identifié que différents neurones n'ont pas tous la même sensibilité à l'incertitude. Par conséquent, il est possible de "lire" l'incertitude contenue dans une image, à partir de l'activité de populations neuronales. Cette "lecture", que nous avons faite grâce à l'Intelligence Artificielle, pourrait également être faite par le cerveau en utilisant ses réseaux neuronaux spécialisés. Ceci expliquerait pourquoi notre perception est si sensible à l'incertitude, et comment celle-ci pourrait nous aider à adapter nos comportements. Malgré le fait que ces neurones sensibles à l'incertitude ne représentent qu'un tiers de la population totale, ils jouent néanmoins un rôle crucial en fournissant à l'ensemble du système neuronal des données qui améliorent le traitement du signal, en y ajoutant une évaluation de sa probabilité et de son degré de confiance.

235
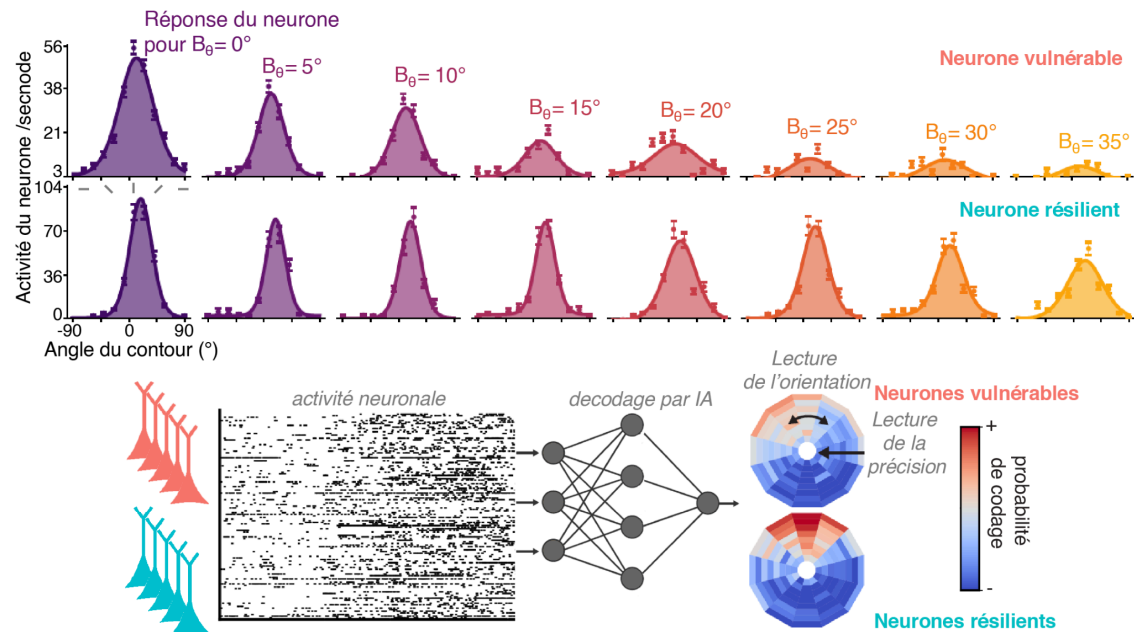
L'activité des neurones est sélective pour certaines orientations de contours, mais peut se révéler soit vulnérable, soit résiliente face à l'augmentation de l'incertitude. En utilisant l'intelligence artificielle, il est possible de déchiffrer le message neuronal, notamment celui qui encode l'incertitude des orientations dans les neurones résilients. La représentation en colonne montre comment l'IA décode l'orientation et l'incertitude d'une image, en représentant des coordonnées polaires (angle et rayon) pour représenter les différentes informations possibles.

Bien que cette découverte soit captivante, elle n'est guère surprenante. Le cerveau est en effet maître dans l'art de naviguer à travers un océan d'informations fragmentées et ambiguës pour forger, grâce à notre perception, une vision cohérente de la réalité. Nous réussissons avec aisance à reconstituer des images même lorsqu'elles sont partiellement voilées par des ombres ou des distorsions. Dans de rares cas, à l'instar du chef-d'œuvre de da Vinci, nous nous trouvons confrontés à un véritable labyrinthe d'ambiguïtés où plusieurs interprétations d'un stimulus sont possibles. Que privilégier alors ? Le sourire énigmatique de la Joconde ou son regard pénétrant qui semble nous suivre ? L'incertitude de ces indices est cruciale pour assembler les fragments de notre perception. Le cerveau, en ce sens, agit comme un sculpteur, modelant notre réalité à partir d'hypothèses fondées sur des probabilités, conformément à l'étymologie du terme "fiction" qui implique le modelage ou le façonnage. D'autant plus que ce processus est dynamique et incorpore diverses sources d'information (vision, proprioception, son, …) pour parvenir à une perception unifiée. Cela s'applique même dans des situations complexes, comme le fait de distinguer plusieurs voix dans le brouhaha d'un dîner.
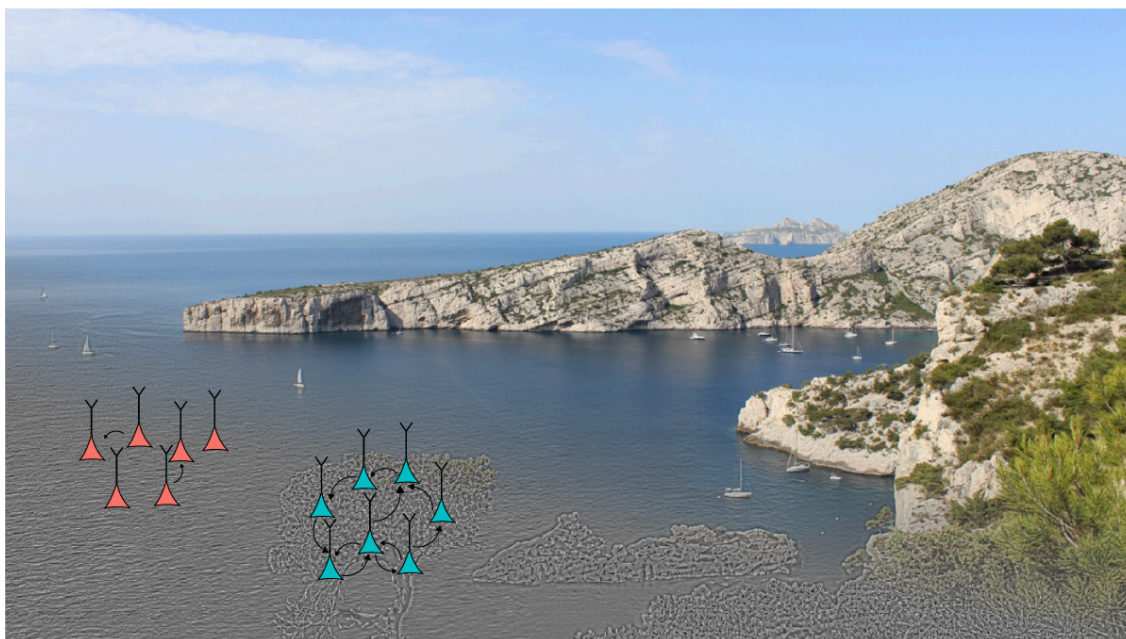
Vers les représentations statistiques jusqu'à une représentation neuronale : les neurones résilients (ici symbolisés en bleu), principalement connectés entre eux, s'opposent dans leurs modes d'interactions par rapport aux neurones vulnérables (en rouge), qui eux se connectent peu.

Ce paradigme a des répercussions importantes pour comprendre divers substrats neuropsychologiques. Des théories récentes confirmées par nos résultats suggèrent que les neurones communiquent entre eux pour échanger des informations sur l'incertitude d'une situation. Par exemple, des déséquilibres dans ce mécanisme d'intégration pourraient expliquer certaines conditions comme dans la schizophrénie, où la perception de la réalité peut être modifiée par une confiance réduite envers le monde externe, jugé trop incertain. Inversement, le traitement privilégié de l'information sensorielle observé dans le spectre des troubles autistiques pourrait être interprété par une balance différente entre un esprit certain de ses perceptions dans un monde toujours empli d'incertitudes.

CERVEAU QUI DOUTE

Nos connaissances et nos incertitudes — qu'elles soient scientifiques, culturelles ou autres — façonnent notre vie et nos sociétés à bien des égards : elles influencent nos comportements d'achat, guident nos choix de vacances, et ont même un impact sur notre appréhension d'enjeux aussi globaux que le changement climatique. Dans ce dernier domaine, chacun de nous joue un rôle, souvent sans en prendre pleinement conscience. Reconnaître les limites de notre savoir et admettre que nous ne détenons pas toutes les réponses permet de mieux maîtriser les éléments du puzzle décisionnel dans lequel nous nous engageons quotidiennement. Ainsi, prendre conscience des mécanismes de notre cognition, qui s'appuie sur des approximations et des incertitudes plutôt que sur une logique rigide, peut nous aider à faire des choix de société plus éclairés et plus humains. Notre monde est de plus en plus influencé par des algorithmes d'intelligence artificielle qui renforcent nos biais cognitifs et cloisonnent nos opinions en des bulles imperméables à l'incertitude. Le passage de l'*Homo Sapiens,* convaincu de ses certitudes, à un *Homo Dubitans*, doué d'un doute rationnel qui reflète son intelligence naturelle, apparaît non seulement souhaitable, mais nécessaire pour faire face aux défis futurs de l'humanité réconciliée avec le vivant.

English translation:
A brain that guesses is a brain that doubts
Introduction

If you were asked which of these two portraits gives the most faithful rendering of a facial expression, you would probably reply that it is the iconic Mona Lisa. However, if you were now asked which of these two faces reveals emotion more clearly and precisely, you would probably find it easier to read the laughing thoughts rendered in Antonello da Messina's portrait. It seems, then, that even if you perceive the Mona Lisa as more "alive", it doesn't make her any more comprehensible to you. Intriguing, isn't it? Although both figures emanate the same air of mystery, there is something about the Mona Lisa that gives it a bewitching singularity.

In the light of modern scientific analysis, we now have a better understanding of why the portrait of Mona Lisa appears so authentic and so mysterious. The work's legendary ambivalence allows the Mona Lisa's gaze to follow the viewer, and her smile to appear to change depending on the angle from which she is viewed. Da Vinci, a keen observer of the world around him, masterfully used realistic techniques to bring to life an image that, while static, evokes a dynamic presence comparable to those we encounter in our everyday lives. The visual ambiguity induced by the observation of the Mona Lisa, a central aspect of its mystery, is reminiscent of the way sunlight finds its way through foliage in a forest, playing subtly with shadows to create an atmosphere that is both lively and enigmatic.

The enigmatic realism of the portrait of the Mona Lisa is down to the talent of its creator. Da Vinci was an undisputed master of sfumato, an arduous artistic technique used to soften the contours of a painting for a more natural, realistic effect. This is a fine example of his multidisciplinary genius, for da Vinci, with only the knowledge of his time, seemed to anticipate certain advances in modern science on natural images. Indeed, we now know that our brains are finely tuned to analyse the visual world around us from every angle. In particular, a system of brain areas processes the information coming from the retina, starting by breaking down the world around us into small, oriented and subtle outline elements, meticulously reproduced in the sfumato of the Mona Lisa. These contours form the framework of our visual perception, from which our brain sketches an image of the luminous world around us.

Analysis of Natural Images

Like the enigmatic smile of the Mona Lisa, natural images are themselves imbued with uncertainty and complexity. Take, for example, the glaring difference between the decompositions of orientations that form the contours of a tree and those that draw a boat, as depicted in the figure above. The different uncertainties of these contours measure fundamentally distinct degrees of ambiguity and perfectly illustrate the variety inherent in our visual environment.

In a modern context, this is all the more important. Imagine yourself as a pedestrian, in the middle of a city, about to cross the road at a crossing without traffic lights. You stare at the driver of the nearest car, trying to read his intentions on his face. Should you trust his staring eyes? Or take into account the uncertainty of his car, which has slowed down but not quite braked? Should you expect a hand gesture to resolve this

uncertainty? So many visual variables to consider in order to reduce the uncertainty as to whether the driver is going to stop, while hoping that his face is less mysterious than the Mona Lisa.

Brain that guesses

So there's no doubt that uncertainty is a fundamental issue in psychology. But what about the 'brain' itself? A recent study we conducted at the Institut des Neurosciences de la Timone (Aix-Marseille University; CNRS) in collaboration with the University of Montreal sheds some light on this subject.

By observing neurons involved in vision, and in particular in the representation of oriented contours, we have identified that different neurons do not all have the same sensitivity to uncertainty. As a result, it is possible to 'read' the uncertainty contained in an image from the activity of neuronal populations. This 'reading', which we have done with Artificial Intelligence, could also be done by the brain using its specialized neural networks. This would explain why our perception is so sensitive to uncertainty, and how uncertainty could control our behaviour. Despite the fact that these neurons sensitive to uncertainty only represent a third of the total population, they nevertheless play a crucial role in providing the entire neural system with data that improves signal processing, by adding an assessment of its probability and degree of confidence.

Although this discovery is captivating, it is hardly surprising. The brain has mastered the art of navigating through an ocean of fragmented and ambiguous information to forge, through our perception, a coherent vision of reality. We can easily reconstruct images even when they are partially obscured by shadows or distortions. In rare cases, such as da Vinci's masterpiece, we find ourselves faced with a veritable labyrinth of ambiguities where several interpretations of a stimulus are possible. So which should we choose? The Mona Lisa's enigmatic smile or her penetrating gaze that seems to follow us? The uncertainty of these clues is crucial in piecing together the fragments of our perception. The brain, in this sense, acts like a sculptor, modelling our reality from hypotheses based on probabilities, in accordance with the etymology of the term "fiction", which implies modelling or shaping. All the more so because this process is dynamic and incorporates various sources of information (vision, proprioception, sound, etc.) to achieve a unified perception. This applies even in complex situations, such as distinguishing several voices in the din of a dinner party.

This paradigm has important implications for understanding various neuropsychological substrates. Recent theories confirmed by our results suggest that neurons communicate with each other to exchange information about the uncertainty of a situation. For example, different balances in this integration mechanism could explain certain conditions such as schizophrenia, where the perception of reality can be modified by reduced beliefs about the external world, which is considered too uncertain. Conversely, the privileged processing of sensory information observed in the autistic spectrum could be interpreted by a different balance between a mind that is certain of its perceptions in a world that is still full of uncertainties.

Brain that doubts

Our knowledge and uncertainties - be they scientific, cultural or other - shape our

lives and our societies in many ways: they influence our purchasing behaviour, guide our holiday choices, and even have an impact on issues as global as climate change. In this last area, each of us plays a role, often without fully realizing it. Acknowledging the limits of our knowledge and admitting that we don't have all the answers means we can better weigh up the pieces of the decision-making jigsaw in which we engage on a daily basis. In this way, becoming aware of the mechanisms of our cognition, which is based on approximations and uncertainties rather than on rigid, infallible logic, can help us to make more enlightened and more humane choices for society. Our world is gradually being governed more and more by artificial intelligence algorithms that reinforce our cognitive biases and compartmentalize our opinions into bubbles impervious to uncertainty. The transition from *Homo Sapiens*, convinced of his certainties, to *Homo Dubitans*, gifted with the rational doubt that reflects his natural intelligence, seems not only desirable but necessary if we are to meet the future challenges facing humanity.