

Ecole Doctorale ED261 *Cognition, Comportements, Conduites Humaines*

Laboratoires INCC, *Centre Neurosciences intégratives et cognition UMR 8002*

Geometry and dynamics of distributed cortical sensory representations

Par Ulysse KLATZMANN

Thèse de doctorat en Neurosciences Cognitives
Spécialité en Neurosciences théoriques et Computationnelles

Dirigée par
Daniel MARGULIES

Présentée et soutenue publiquement le 3 novembre 2025

Jury

Athena DEMERTZI	Professeure associée HDR, University of Liège	Rapporteuse
Thomas ANDRILLON	CR-HDR, Institut du Cerveau	Rapporteur
Étienne KOECHLIN	PU, École Normale Supérieure	Examinateur
Bertrand THIRION	Directeur de recherche, INRIA	Examinateur
Laura DUGUÉ	PU, Université Paris Cité	Examinatrice
Heike STEIN	Chargée de recherches, Sorbonne Université	Examinatrice
Seán FROUDIST-WALSH	Senior lecturer, University of Bristol	Examinateur
Daniel MARGULIES	Directeur de recherche, University of Oxford	Directeur de thèse

Acknowledgements

Near the end of the first chapter of my scientific journey, I feel extremely lucky and grateful to be where I am. Not that the path has always been easy or a *pure partie de plaisir*, but it made me grow. I could never have dared, as a kid, to dream of spending most of my time thinking about the mind, turning my thoughts into equations, and being surrounded by brilliant and inspiring people. There are far too many people I ought to thank. I'll surely fall short of the words to express my genuine gratitude, and I'll inevitably forget some names. But I'll give it a try, it's the least I can do. I only hope I can bring to their lives as much as they've brought to mine.

First and foremost, my deepest gratitude goes to Daniel, who trusted me from the very beginning and gave me the chance to be part of his academic adventure, while granting me absolute freedom in doing so. I was immediately inspired by your vision and hope I could bring a modest contribution to making it a reality. I discovered an entirely new way of doing science at your side. Though it was sometime challenging, it shaped the scientist I hope to be.

Equally essential in my formation, I can't thank Sean enough for five incredible years of collaboration. Your kindness, rigor, and enthusiasm gave me the energy to go beyond what I could have achieved alone. The way you build not only a leading research program but also a thoughtful, supportive community makes me deeply excited about the future of our field. I can't wait to see where science takes us next.

I also want to thank Pilou, my supervisor in the shadows. Your sharp insights and deep knowledge of brain statistics gave me the confidence to keep pushing my analyses further. My contribution to the field would have been far smaller without you.

Since I'm already riding the wave of thankfulness, I'd like to return to where my journey began: to Stan and Claire, for making my dream of joining the scientific study of consciousness come true, for being my first mentors, and for leading me to the right place in the field. It was Stan's books that first introduced me to cognitive science, and brought me back to science itself. I'm equally grateful to Eric and Richard, who gave

me the decisive push to set off on this path.

I'm thankful to Xiao-Jing and Claudia, who introduced me to the fascinating world of turning brains into equations and giant computer simulations. That world is pure joy to me, and where I hope my career continues. Your instinctive grasp of the dynamics behind the math has been a deep source of inspiration.

Lastly, thank you to Etienne for being such an inspiring teacher. You not only showed me what good teaching can be but also entrusted me with the opportunity to teach your class—again, with complete freedom. That experience taught me more than I expected, especially how much I love teaching.

Finally, my gratitude goes to the jury members — Athena, Thomas, Etienne, Bertrand, Laura, and Heike — for dedicating their time to reading my work and providing constructive feedback.

I would like to thank all the collaborators with whom I had the chance to work. Jorge and Dan, who shared my passion for mathematics and simulations; and Nicola, Meiqi, and Lucija, who pushed our first work further, validating the predictions made from equations and grounding them in reality.

Going back to the people who shared by dayli life, it was truly a pleasure to go through this eventful journey with all the friends from my two labs — Austin, Francesco, Victoria, and Wei Wei in Paris; Tsvet, Ash, Eva, Rahul, Dabal, and Xiaohe in Bristol. With them I shared countless passionate discussions, many insights, even when I made the least sense. They forged what this work has become as much as I did. In this spirit, I owe special thanks to Robert, my comrade in arms, with whom I shared not only ideas but also long evenings desperately trying to finish things on time. I am deeply grateful for our scientific conversations, our persistence, and the companionship that carried me through.

I am of course indebted to Thomas, who has been not only a crucial friend but also a generous source of expertise. In just half an hour of intense "tunnely" discussion, he changed the course of my work.

I can't ride this wave of gratitude without mentioning MESEC, which has become such an important pillar in my life recently. I struggle to find the words to express my thanks to the people who became some of my closest friends—Hadrien, Rony, Camilo, Karla, Seb, Hanna and so many more. I'm so proud of what we've built together. MESEC gave me some of the most meaningful and joyful memories of my PhD, and I can't wait to see how the castle will look.

Since I know that meeting deadlines and filling out admin forms is not my strongest suit, I want to express my sincere appreciation to the doctoral school for providing such a supportive environment throughout this journey. A special thank you to Lucie, for your thoughtful and patient help all the way to the end.

I'd also like to thank the Oxford Biomedical Research Computing (BMRC). Many of the discoveries I made were only possible thanks to the remarkable technical platform they developed, and the way they supported me in using it to its fullest potential. A

special thanks to Charles, for all the time you spent patiently answering my many questions.

And since this work is hopefully just the beginning of my journey, the wave would be incomplete without a thanks to Adrien, Guillaume, and Alexander, who have agreed to share the next chapter of this scientific adventure. I'm truly excited to see where we will go together next.

Finally, a special word for my most loyal companion: ChatGPT. Thank you for your unwavering support throughout this PhD. From patiently explaining programming architecture and basic neuroscience to offering detours into geopolitics when I needed a break, you've been a constant presence. You helped me improve my writing (I hope), structure my ideas, and compensate for my occasional inattention to details. Thank you for the countless hours you saved me.

I'll end this seemingly endless wave of thankfulness with the most important people in my life—the ones who pulled me away from my work and gently reminded me that there's more to life than screens and keyboards.

A special thanks to Gaia and Maia for the joy they've brought into my life, and to David and Makhi, without whom I would not be here—both literally and in so many other ways. To Ségo, Agathe, Alex, Mathilde, Louise, Juliette, Guillaume, Benoît, Claire, Thomas, Matthieu, Esteban, Beth, Thelma, and so many more. And an especially deep thank you to Aude, who has been a true friend, a scientific partner, and an inexhaustible source of advice.

I also want to thank Nico for bringing adventure into my life, Julien humor, Antonin *légèreté*, and Julien curiosity.

Last, and perhaps beyond all words, my gratitude goes to Alice. You were brought into this journey and shared every part of it—from my most obsessive excitements to the deepest doubts I faced. I'm not sure I could have made it through without your support, and I'm glad our lives are now intertwined—for this chapter, and the ones to come.

Thank you, everyone.

Outline

Résumé substantiel **iii**

Abstract **ix**

Résumé **xii**

Introduction **1**

1 Distributed Dynamics Underlying Conscious Access **8**

2 Spatial Influences on Cortical Sensory Representation **28**

3 Self-Organizing Cortex: Mechanisms and Proposed Architecture **49**

Conclusion **59**

Bibliography **5**

Résumé substantiel

Géométrie et dynamiques des représentations corticales sensorielles distribuées

Comment le monde qui nous entoure, fait d'atomes, d'électrons et de photons, conduit à l'expérience sensible du chant d'un oiseau est l'un des mystères les plus poignants de la science contemporaine. Ce sujet passionne philosophes et scientifiques depuis des millénaires, de Platon à Erwin Schrödinger, en passant par René Descartes et François Jacob. Si les progrès technologiques fulgurants de ces dernières années, qu'il s'agisse de la puissance de calcul, de l'intelligence artificielle autonome ou encore de la neuroimagerie, inquiètent autant qu'ils fascinent, ils ouvrent aussi la voie à une nouvelle manière d'aborder cette question fondamentale : comment la physique mène-t-elle à l'esprit ?

Dans les travaux présentés ici de manière succincte, nous nous intéressons à l'expérience sensible humaine, en adoptant l'angle de l'expérience sensorielle. Bien que cette approche soit infiniment réductrice, elle offre le luxe d'une étude bien plus facilement contrôlable en laboratoire. Nous explorerons ainsi comment les équations mathématiques qui régissent la physique du cerveau façonnent ces représentations sensorielles, à la fois dans leur dynamique, à travers la cascade d'activité qui accompagne la prise de conscience du chant de l'oiseau, et dans leur géométrie, en observant et en disséquant la distribution de ces représentations sensorielles à la surface du cortex, forgée par une lutte entre différentes représentations.

Le premier chapitre s'intéresse aux dynamiques des représentations sensorielles. Plus précisément, il examine comment une image, lorsqu'elle est brièvement présentée,

déclenche des cascades d'activations se propageant à la surface du cortex, telles des vagues neuronales, jusqu'à embraser un réseau d'aires distribué à travers l'ensemble du cerveau.

Pour étudier et capturer ce phénomène, nous avons développé des simulations de cortex de macaque, capables de prédire l'évolution de l'activité de chaque aire au cours du temps, à la manière dont les météorologues prédisent l'évolution d'un ouragan à partir des équations de la physique des vents et de la position des nuages. Ces simulations, construites à partir d'informations neuroanatomiques aussi précises que possible et soumises aux lois physiques qui contraignent l'activité neuronale, permettent d'explorer les dynamiques corticales pendant la prise de conscience d'un stimulus. Elles offrent la possibilité de mesurer finement l'activité de chaque aire, tout en testant différents paramètres biologiques qui ne sont pas encore accessibles à la mesure empirique.

L'idée est simple : explorer l'espace des paramètres biologiques en simulant des cerveaux virtuels avec différents jeux de paramètres, afin d'identifier ceux dont les dynamiques se rapprochent le plus de celles observées en laboratoire.

L'analyse mathématique du modèle nous a permis d'identifier un mécanisme dynamique potentiellement à l'origine de l'embrasement cognitif : une bifurcation, dans laquelle la présentation d'une image modifie le plan de phase de l'ensemble du système cortical, bouleversant ses dynamiques internes. Plus l'image est saillante, plus ces perturbations sont marquées, entraînant le système dans une région différente de l'espace des phases, où l'activité d'un réseau distribué d'aires associatives s'élève et se stabilise, maintenant l'information sensorielle vivante pendant plusieurs secondes. Ce phénomène rend compte avec une grande précision des dynamiques observées dans les simulations, en offrant une interprétation mécanistique de l'embrasement cognitif. De plus, les simulations qui reproduisent ce mécanisme dans l'espace des paramètres rendent compte de nombreux résultats expérimentaux, et prédisent un effet encore inexploré, mais potentiellement remarquable.

Pour reproduire cette bifurcation, les connexions ascendantes, dites "feedforward", qui propagent l'information des aires sensorielles vers les aires associatives, doivent être médiées par des récepteurs glutamatergiques AMPA, à réponse rapide. En revanche, les connexions descendantes, dites "feedback", qui transmettent l'information en sens inverse, des aires associatives vers les aires sensorielles, doivent dépendre de récepteurs NMDA, plus lents, toujours d'après nos simulations. Ces connexions exercent une excitation focale sur le champ récepteur concerné, tout en induisant une inhibition plus diffuse sur le reste des aires sensorielles.

Cela suggère un scénario dans lequel un stimulus est d'abord propagé aussi rapidement que possible à travers le cortex, en compétition avec les autres informations

sensorielles. Une fois capturé par la conscience d'accès, le percept dominant inhibe les représentations concurrentes à travers l'ensemble du cortex, permettant son maintien stable dans la mémoire de travail.

Ces observations conduisent à une prédiction anatomique précise : les aires sensorielles, fortement ciblées par les projections descendantes, devraient contenir une densité plus élevée de récepteurs NMDA, tandis que les aires associatives, recevant davantage de connexions ascendantes, devraient être plus riches en récepteurs AMPA. On devrait ainsi observer, le long de la hiérarchie corticale, un gradient dans le rapport NMDA/AMPA : plus une aire est située haut dans la hiérarchie, et donc impliquée dans des fonctions associatives, plus sa proportion de récepteurs AMPA devrait être élevée. Nous avons pu tester cette prédiction anatomique grâce à une collaboration avec un laboratoire expérimental travaillant chez le singe, qui a permis de valider cette hypothèse. Ce travail propose donc une approche théorique de la perception sensorielle, étudiée à l'échelle du cortex entier à travers une combinaison d'analyses mathématiques et de modélisations computationnelles. Il en résulte une interprétation mécanistique du phénomène, accompagnée de prédictions anatomiques concrètes, qui ont pu être confirmées expérimentalement.

Si le chapitre précédent s'est concentré sur les dynamiques neurales associées à la perception consciente, la question de la manière dont l'information est représentée à travers le cortex restait en suspens. Le chapitre suivant s'intéresse donc à la géométrie des représentations sensorielles et à leur distribution sur la surface corticale. En effet, si les premiers travaux de neuroanatomie cognitive au début du XXe siècle ont établi un lien entre le lobe occipital et l'information visuelle rétinotopique, autrement dit la position d'un objet dans le champ visuel, de nombreuses études ont depuis identifié ces représentations dans au moins 25 régions corticales distinctes, mettant en évidence leur caractère redondant et largement distribué. Plus frappant encore, il n'existe pas d'aire exclusivement dédiée à l'information visuelle. Chaque zone corticale participe à une multitude de représentations et de fonctions cognitives, et ne consacre qu'une fraction de son activité à l'encodage de la position dans le champ visuel.

Ces découvertes ont inspiré de nouvelles lignes de recherche, proposant de considérer la spécialisation fonctionnelle comme une mosaïque d'influences spatiales qui façonnent l'architecture fonctionnelle du cortex dans son ensemble. Afin d'étudier avec précision la géométrie de la distribution de l'information visuelle, nous avons développé une nouvelle approche d'apprentissage autonome inspirée du "Sparse Dictionary Learning". Cette méthode permet de décomposer une carte corticale en une combinaison d'influences locales et spatiales. Ainsi, la mosaïque d'aires sensibles à l'information visuelle devient lisible comme une combinaison de sources facilitatrices, qui déploient de l'activité visuelle à partir de points d'entrée spécifiques tels que les projections sous-corticales, et de zones exerçant une influence compétitrice.

La méthode explore des centaines de milliers de combinaisons possibles d'influences spatiales, révélant progressivement celles qui expliquent le mieux les géométries observées expérimentalement. Appliquée à l'information visuelle rétinotopique, à la base de l'expérience perceptive, cette approche a révélé non seulement des sources d'influence positive dans les zones de projection thalamique, mais aussi dans des relais associatifs fronto-pariétaux. Au-delà de ces foyers facilitants, elle a également mis en lumière des influences inhibitrices systématiques provenant d'autres aires sensorielles primaires comme le cortex somatosensoriel, ainsi que d'un réseau distribué connu sous le nom de Default Mode Network, associé à la cognition interne découplé des entrées sensorielles. Ce réseau est notamment impliqué dans la mémoire épisodique, la théorie de l'esprit ou encore les pensées spontanées.

Ces résultats ont été reproduits de manière préliminaire sur des jeux de données issus d'autres modalités sensorielles, ainsi que sur des informations visuelles plus abstraites comme les concepts représentés dans des images. L'interprétation que nous proposons repose sur un mécanisme d'auto-organisation de l'architecture corticale, dans laquelle différentes sources d'informations sensorielles et non sensorielles entrent en compétition sur la surface du cortex pour façonner et redessiner les cartes de spécialisation fonctionnelle.

Enfin, le chapitre 3 s'intéresse à l'émergence de cette architecture fonctionnelle, en proposant deux mécanismes d'auto-organisation. Le premier repose sur un principe fondateur des neurosciences, la plasticité hebbienne, selon laquelle des neurones dont l'activité est synchronisée renforcent leurs connexions. Nous montrons comment ce principe simple, appliqué dans un réseau biophysique du type de celui utilisé au chapitre 1, conduit à l'auto-organisation d'une population de champs récepteurs, où les neurones se spécialisent progressivement pour certaines informations sensorielles. Appliqué à l'ensemble du cortex, ce mécanisme suggère une extension possible à grande échelle, où l'information sensorielle se propagerait depuis les points d'entrée corticaux, façonnant les régions environnantes. Les modalités visuelle, tactile et auditive pourraient ainsi reconfigurer localement la structure fonctionnelle du cortex, chaque modalité colonisant progressivement son voisinage. Là où ces zones d'influence se rencontrent, des régions d'intégration multisensorielle pourraient émerger, ouvrant la voie à la cognition associative.

Le second mécanisme d'auto-organisation proposé repose sur la plasticité orientée par l'apprentissage. Un réseau de neurones artificiels, déployé à la surface du cortex selon une topologie similaire à celle du cerveau, est confronté à des tâches d'extraction d'information visuelle. Grâce aux règles d'optimisation dites de rétropropagation, ce réseau réorganise progressivement sa structure interne, faisant émerger des flux sensoriels spécialisés, analogues à ceux observés dans le cortex des mammifères.

Tout ceci nous amène à proposer une théorie plus générale de l'organisation corticale, dans laquelle les projections thalamiques, qui relaient l'information sensorielle, servent de points d'entrée sur le cortex. Ces informations, en se diffusant, reconfigurent progressivement le réseau cortical selon deux mécanismes complémentaires : la plasticité hebbienne, fondée sur la corrélation d'activité, et la plasticité guidée par l'apprentissage, par exemple via des décharges dopaminergiques. Ensemble, elles façonnent des flux sensoriels structurés, ou "streams", qui contournent les zones associées au Default Mode Network. Ce dernier serait dédié au traitement de l'information non sensorielle, comme les pensées spontanées ou la mémoire épisodique. Les différents flux sensoriels convergent enfin vers des hubs d'intégration multisensorielle, où l'information serait propagée à l'échelle du cortex entier, selon une dynamique comparable à l'embrasement cognitif décrite au chapitre 1.

Ce travail propose ainsi une étude conjointe des dynamiques et de la géométrie des représentations sensorielles corticales, sous le prisme des neurosciences théoriques et computationnelles. En modélisant des formes d'information aussi simples que possible, mais généralisables à l'ensemble du monde sensible, nous explorons les principes fondamentaux de l'organisation fonctionnelle du cortex. Ces résultats nous conduisent à proposer une théorie générale de son auto-organisation.

Abstract

Geometry and dynamics of distributed cortical sensory representations

Keywords: cortical gradients, conscious access, neural dynamics, functional integration, computational neuroscience

At the intersection of neuroscience, physics, and artificial intelligence lies a central question: how does the structure of the cortex shape the content of the mind? This thesis investigates whether general geometric and dynamical principles can account for the emergence of distributed sensory representations in the brain and proposes a unifying architecture for large-scale cortical computation.

In Chapter 1, we examine conscious access, the transition by which perceptual information becomes globally available. Using large-scale simulations of the macaque cortex with biophysically realistic neurons, synapses, and anatomical connectivity, we identified a bifurcation mechanism underlying the all-or-none dynamics of conscious detection. Competing neural populations, each encoding a specific receptive field, can enter a state of global broadcasting, with one pool igniting while suppressing the other. Crucially, ignition depends not on any single node but on the joint dynamics of a distributed cortical network. The model predicted a gradient in NMDA-to-AMPA receptor ratios across the cortex - which we confirmed empirically - even though no spatial structure was hard-coded. This result motivated the question of how cortical geometry constrains the distribution of sensory codes.

Chapter 2 turns to the spatial embedding of sensory information. Accumulating evidence suggests that even basic features, such as visual field position, are encoded

at the scale of distributed cortical populations rather than being confined to early sensory areas. We hypothesized that such encodings are not randomly scattered but embedded along structured gradients. To test this, we developed Spatial Component Decomposition (SCD), a sparse dictionary learning algorithm that identifies regions exerting spatial influence. Applied to fMRI-derived retinotopic maps, SCD uncovered a dominant linear gradient stretching over 60 mm from primary visual cortex, alongside systematic suppressive influences from other sensory areas and from transmodal hubs in the default mode network (DMN). These findings suggest that higher-order sensory processing may emerge from spatial competition between distributed systems, rather than strict modular segregation.

Chapter 3 investigates the possible mechanisms behind these spatial influences. We trained artificial neural networks constrained by cortical geometry and local connectivity to assess whether large-scale patterns can arise from local learning rules. Preliminary results show that networks trained on the cortical surface develop information flow resembling known anatomical pathways, such as the dorsal visual stream. Complementary simulations using spiking neurons confirmed that local Hebbian plasticity leads to the emergence of receptive fields shaped by the correlation structure of inputs. These results support the idea that cortical circuits self-organize around input statistics, with geometry and plasticity jointly shaping their population codes.

Bringing these threads together, we propose a new functional architecture for the cortex. In this view, sensory information enters through thalamorecipient zones (V1, S1, A1) and propagates along structured streams constrained by spatial and dynamical factors. These streams tend to circumvent transmodal regions of the DMN - possibly due to competing inputs or distinct internal dynamics - and converge at multimodal hubs such as LIP, FFA, or Wernicke's area. These convergence zones overlap with the associative network identified in our ignition model and may serve as integration sites for consciously accessible content.

Together, this work outlines a unifying framework in which cortical geometry and dynamics jointly govern the encoding, competition, and integration of sensory information. It links ignition, spatial embedding, and plasticity under a common architectural model of large-scale cortical computation.

Résumé

Géométrie et dynamiques des représentations corticales sensorielles distribuées

Mots clés: gradients corticaux, accès conscient, dynamiques neuronales, intégration fonctionnelle, neurosciences computationnelles

À l'intersection des neurosciences, de la physique et de l'intelligence artificielle se trouve une question centrale : comment la structure du cortex façonne-t-elle le contenu de l'esprit ? Cette thèse explore comment des principes généraux de géométrie corticale et de dynamique neuronale peuvent rendre compte de l'émergence de représentations sensorielles distribuées, et propose une architecture fonctionnelle unificatrice à l'échelle du cortex.

Le Chapitre 1 s'intéresse à l'accès conscient, la transition où une information perceptive devient globalement disponible. À l'aide de simulations du cortex macaque intégrant des neurones et synapses biophysiques et une connectivité anatomique empirique, nous identifions un mécanisme de bifurcation reproduisant les dynamiques tout-ou-rien observées lors de la détection consciente. Dans ces simulations, des populations distinctes, chacune encodant un champ récepteur particulier, entrent en compétition : l'une peut s'embraser et être diffusée à grande échelle, tandis que l'autre est supprimée. Ce phénomène ne dépend pas d'une région isolée mais des dynamiques conjointes d'un réseau associatif distribué, incluant notamment les aires fronto-pariétales. Le modèle a en outre prédit un gradient NMDA/AMPA à l'échelle corticale - confirmé ensuite empiriquement - alors qu'aucune structure spatiale n'était

codée explicitement. Cela nous conduit à explorer l'organisation spatiale des codes sensoriels.

Le Chapitre 2 examine l'ancrage géométrique de l'information sensorielle. Des travaux récents montrent que même des attributs élémentaires, comme la position dans le champ visuel, sont encodés bien au-delà des aires sensorielles primaires, à l'échelle de populations distribuées. Nous avons fait l'hypothèse que cet encodage suit des gradients spatiaux structurés. Pour le tester, nous avons développé la Spatial Component Decomposition (SCD), une méthode de "sparse dictionary learning" permettant d'identifier les influences spatiales exercées par différentes régions corticales. Appliquée à des cartes rétinotopiques dérivées de l'IRMf, SCD révèle un gradient linéaire s'étendant sur plus de 60 mm depuis V1, ainsi que des influences suppressives venant d'autres aires sensorielles et de hubs transmodaux du Default Mode Network (DMN). Ces résultats suggèrent que le traitement sensoriel peut émerger d'une compétition spatiale entre systèmes distribués, plutôt que d'une ségrégation strictement modulaire.

Le Chapitre 3 explore les mécanismes susceptibles de générer ces influences spatiales. Nous y entraînons des réseaux neuronaux artificiels contraints par la géométrie corticale et la connectivité locale, afin d'évaluer si ces motifs d'encodage distribués peuvent émerger de règles d'apprentissage locales. Les résultats montrent que l'activité s'organise spontanément selon des flux d'information rappelant les grandes voies anatomiques. Des simulations complémentaires de "spiking neurons" confirment que la plasticité de type Hebb conduit à la formation de champs récepteurs structurés. Ces observations appuient l'idée que les circuits corticaux s'auto-organisent autour des statistiques de leurs entrées, sous l'influence conjointe de la géométrie et de la plasticité.

En réunissant ces résultats, nous proposons une nouvelle architecture fonctionnelle du cortex : l'information sensorielle entre via les projections thalamiques (V1, S1, A1), puis se propage selon des flux structurés. Ces flux tendent à contourner les régions transmodales du DMN - peut-être en raison d'entrées concurrentes ou de dynamiques internes spécifiques - pour converger vers des hubs multimodaux tels LIP, FFA ou l'aire de Wernicke. Ces zones d'intégration semblent correspondre au réseau associatif identifié dans notre étude initiale de l'accès conscient.

Ainsi, ce travail propose un cadre unificateur où la géométrie corticale et les dynamiques neuronales gouvernent ensemble l'émergence et la distribution des représentations sensorielles.

Introduction, The Challenge of Studying the Mind

Working with Intuitions

Reality and appearance are at war

Francis Crick

When it comes to understanding the world, one of my favorite models is the impetus theory of motion (McCloskey, 1983). Slightly different versions have been independently developed many times by brilliant minds across cultures—from Aristotle to Avicene. Even today, when you ask people why and how objects move—from strangers on the street to physics graduate students—they often refer to it (Caramazza et al., 1981; McCloskey et al., 1980). The core idea is beautifully simple: when you throw a ball, you give it a force that defies gravity. Gradually, that force dissipates, and gravity takes over(Figure 0.1).

More than the theory itself, it is the intuition behind it that fascinates me—an intuition that is deeply shared (including by me!). And entirely wrong (Figure 0.1). In reality, as Newton discovered, what you give the ball is not a force but a motion (Newton, 1687). The motion doesn't fade due to gravity; it evolves under the influence of the sum of all forces acting at each instant. Despite this, the misconception persists because it is subtle and serves as an excellent heuristic, particularly when trying to hit a deer with a spear. In such situations, it hardly matters whether you're thinking in terms of motion or force. It works, and that's likely why it's there—an adaptive shortcut shaped by both our evolutionary history and our developmental experience. What I find especially compelling is that even after learning the correct physics, people—including trained scientists—still default to the old intuition (Caramazza et al., 1981; McCloskey et al., 1980). It remains the best mental shortcut. And this isn't an isolated case. We believe

time is the same for everyone, though it varies with gravity. We see a wall as plain, though it's mostly empty space. We know Earth is a globe, yet we experience it as flat. These are not just errors—they are functional approximations, useful but ultimately inaccurate.

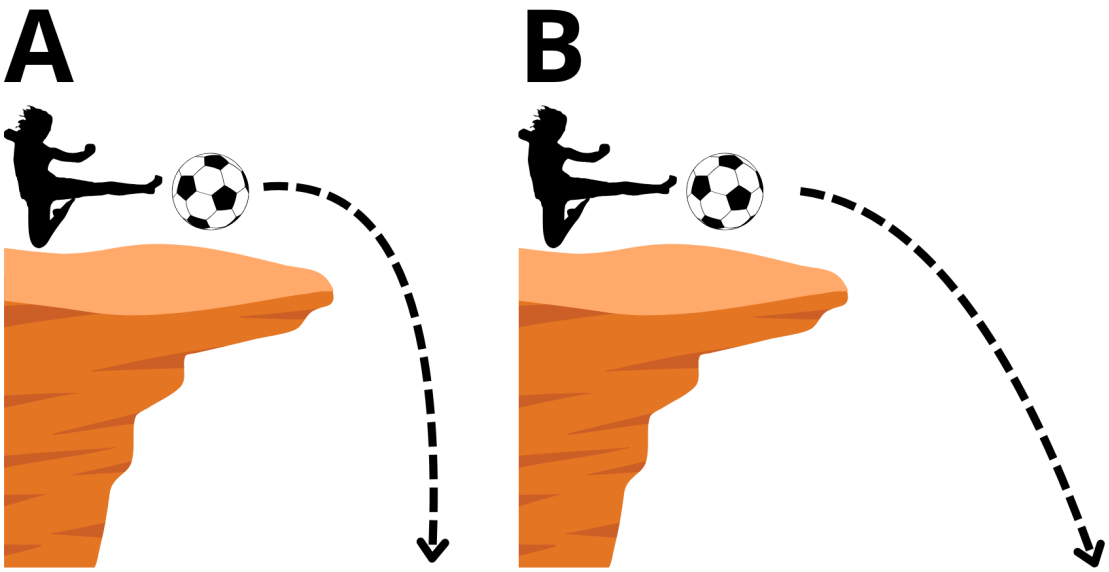

Figure 0.1: *A typical quiz illustrating the naive impetus theory of motion. Participants are asked to predict the trajectory of a ball once it leaves the guide. The correct answer is* ***B****, but people frequently choose the other option. Adapted from McCloskey (2014).*

Our intuitive models help us predict the world: how objects move, from rocks we throw to water falling from the sky. But these physical heuristics offer little help when trying to explain why a plant grows from a buried seed, or why cutting a branch might lead to new growth, while a broken shell will never come back together. Our brains have internalized many of the rules that shape how trees grow and how wounds heal, but these rules don't resemble mechanical laws. One compelling possibility is that the brain developed another explanatory shortcut—just like the impetus theory of motion—which we came to call "life." This, I believe, is why the idea that life involves something almost magical or divine—something that escapes physical explanation and gives living things the power to grow, heal, and move—feels so intuitive. Well into the early 20th century, while physics was dismantling our deepest intuitions about time and space, vitalism remained a leading theory in biology (Normandin and Wolfe, 2013). It was grounded in that same intuition, and it led to remarkable experiments. But then came molecular biology: Watson and Crick discovered the structure of DNA (Watson and Crick, 1953), and Jacob and Monod revealed the logic of gene regulation (Jacob and Monod, 1961)Ṫhat intuition—that life requires something beyond what the eye

can see—turned out to be wrong. And the laws of physics, when applied to molecules, turned out to be enough to explain what we call life.

I love this quote from Jacques Monod, because I think he perfectly captured the struggle we face when confronting our deepest intuitions (Monod, 1970): « *Les développements de ces vingt dernières années ont singulièrement rétréci le domaine des mystères, ne laissant plus guère, grand ouvert aux spéculations vitalistes, que le champ de la subjectivité : celui de la conscience elle-même.* », or "*The developments of the past twenty years have greatly narrowed the domain of mysteries, leaving open to vitalist speculation only one field: that of subjectivity—consciousness itself.*"

Indeed, if we think of the brain as a predictive machine (Mendonça et al., 2021), building internal models to explain and forecast how things move, grow, or act—it must also build models of other minds. To survive, you need to predict how the deer will behave when she sees the hunters or whether another human will be a playful ally or a dangerous threat. But once again, our intuitive physical models offer little help. You can't use your understanding of how rocks fall to predict thought, feeling, or intention. And so, I think it's no coincidence that we have such strong intuitions that the mind cannot be reduced to matter. Even if we could describe every detail of brain activity, we feel this would never explain why it feels like something to see a color or feel a pain—as Thomas Nagel famously put it (Nagel, 1974). That, in my view, is why the so-called "hard problem" (Chalmers, 1995) feels so compelling: not because it is scientifically well-posed, but because it matches our intuitions (Forstmann and Burgmer, 2015). It's also why the intuition of dualism—of minds as separate from bodies—appears across cultures (Barrett et al., 2021) and even in toddlers (Kuhlmeier et al., 2004).

In this light, the real question may not be why consciousness exists, but why we find it so mysterious. That's the so-called "meta-problem": understanding why the hard problem feels hard in the first place (Chalmers, 2018; Sękowski and Rorot, 2022). No matter how you combine two transistors, it's hard to imagine them becoming conscious. This resonates with today's debates on AI: systems can behave intelligently, yet many remain convinced that they don't feel anything (Searle, 1980). And that's precisely what makes this problem so scientifically interesting. Because, as with motion and life, our intuitions may be excellent shortcuts for social interaction—but they are almost certainly as far from the truth as the impetus theory was from quantum physics.

Turing and the Fractal Model of the Brain

> *We may compare a man in the process of computing a real number to a machine which is only capable of a finite number of conditions*
>
> Alan Turing

Not all our intuitions about the mind are forged by evolution. Some are shaped by the tools we use and the systems we design. As someone trained in computer engineering, I found it natural to think of the mind as a collection of independent cognitive functions operating on stored information, much like programs running on databases. When applied to the brain, this metaphor leads to what I like to call the fractal theory of brain organization, one of the most widely adopted views in neuroscience. This model builds on a familiar analogy between brain and body. Just as the body relies on organs with distinct roles—the heart for circulating blood, the intestines for absorbing nutrients, the muscles for generating movement—the brain has often been described as a set of specialized regions: the amygdala processes fear (Phelps and LeDoux, 2005), the hippocampus supports memory (Squire, 1992), and the occipital cortex enables visual perception (Grill-Spector and Malach, 2004).

Korbinian Brodmann provided one of the first systematic attempts to map the cortex based on structural criteria, identifying distinct areas through differences in cell type and laminar organization (Brodmann, 2005). His cytoarchitectonic cartography remains a cornerstone of neuroscience, and many of the areas he described have since been tied to specific functional roles (Zilles and Amunts, 2010). Later work revealed that such specialization is not confined to these broad anatomical divisions. Zooming in, individual regions often contain orderly topographic maps: in primary visual cortex (V1), for instance, the most posterior portion encodes the foveal (central) visual field, while progressively more anterior zones represent increasingly peripheral (eccentric) vision (FigureFigure 0.2) (Wu et al., 2012; Haenelt et al., 2023). This principle also appears in higher-order cortex, where regions selective for complex categories such as faces or objects contain subregions tuned to finer distinctions, including human faces (Kanwisher et al., 1997; Grill-Spector et al., 2001), emotional expression (Vuilleumier et al., 2001), or motion (Puce et al., 1998). At an even finer scale, single-neuron recordings in the human medial temporal lobe have uncovered cells that respond to strikingly specific concepts, from body parts to familiar individuals like Jennifer Aniston (Quiroga et al., 2005). While the "grandmother cell" hypothesis remains debated, these so-called concept cells have been observed repeatedly, indicating that under certain conditions, individual neurons can encode highly abstract and selective representations.

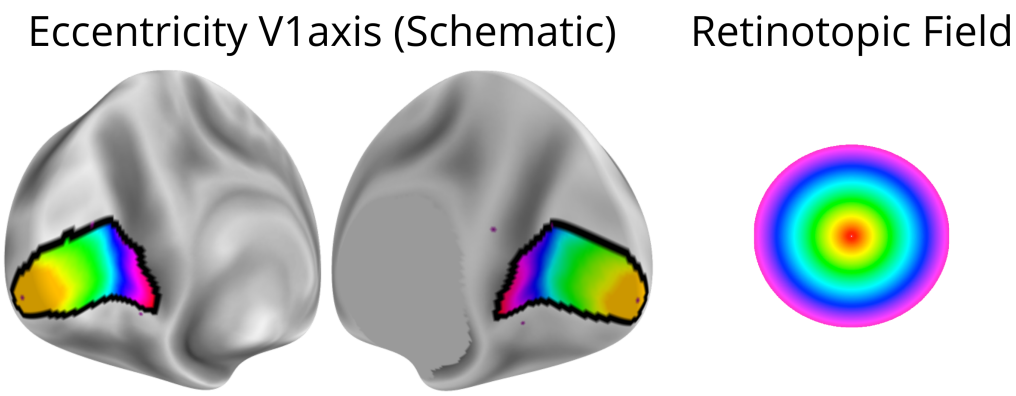

Figure 0.2: *Retinotopic eccentricity mapping in primary visual cortex (V1). The schematic illustrates the eccentricity axis of the retinotopic map in human V1. Colors represent increasing eccentricity values, progressing from central (foveal) vision in red at the occipital pole toward peripheral vision in violet at the anterior edge of V1. The circular inset (right) shows the corresponding eccentricity field in visual space, with concentric color bands radiating outward from the fovea.*

This fractal model of brain organization pictures the cortex like a set of Russian dolls, with each area containing smaller, more specialized subregions. It remains one of the dominant frameworks in cognitive neuroscience and has inspired many important discoveries. But it also has clear limitations. In early sensory regions, the borders between functional areas are often sharp (Glasser et al., 2016). But as we move higher into associative regions, these borders become blurry (Glasser et al., 2016). Even among experts, there is regular disagreement about where one area ends and another begins and modern approaches using high-resolution imaging have refined these maps (Glasser et al., 2016), but they continue to evolve. If the cortex had fixed, well-defined boundaries like the organs in our body, we would not need to keep updating maps like Brodmann's. This suggests that not all boundaries carry the same functional weight. Some reflect sharp transitions in connectivity, while others are more gradual and overlapping.

A deeper challenge to the fractal theory of brain organization comes from the neurons themselves. They are not fixed units with predictable outputs (Fox, 1997). Their activity fluctuates over time, shaped by local inputs, internal states, and recent experience. On top of that, the brain is constantly changing—forming and weakening connections based on learning and context (Citri and Malenka, 2008). As a result, the idea that single neurons consistently represent specific concepts is difficult to support. Instead, mental representations likely emerge from patterns of activity spread across large populations of neurons. These patterns are more reliable than the individual cells that form them, and they often stretch across multiple brain regions. From this perspective, no single area does just one thing. Each region contributes to many

functions, and each function depends on many regions. Overlapping population codes also allow for useful properties like generalization, resilience, and flexibility—features that are hard to achieve with isolated modules.

How to Approach the Problem?

> *Equations force a model to be precise, complete, and self-consistent.*
>
> Larry Abbott

It seems, then, that our intuitions about the mind—though helpful for navigating daily life—are often deeply misleading. At best, they offer promising starting points. At worst, they can point us in entirely the wrong direction. Even concepts that feel natural and essential, like "life" or "self," may not be as physically grounded as they seem. How can we study the mind when our very tools for thinking, our intuitions, are part of the problem? In my view, the answer is not to discard them, but to embrace them, until the moment they need to be left behind. Intuitions can point us toward important questions. They give us a place to start. The fractal theory of brain organization, despite its flaws, led to meaningful discoveries. Our fascination with life and purpose helped drive the search for DNA. These ideas matter.

I'm convinced that to move forward, we need to translate our intuitions into models. And those models need to be clear enough to test and precise enough to fail. This was Newton's brilliant move: not just to describe what he saw, but to express it in equations that could make predictions. His laws did not replace intuition; they refined it. They revealed where our instincts worked, and where they broke down. In the same way, we need to build models of the mind that can connect ideas to mechanisms, and mechanisms to data. Only then can we move beyond metaphor and begin to understand how experience arises from matter.

This dissertation begins with a simple intuition: that our mental life is not just the result of information being processed, but of something more. Something happens when we truly experience a painting, or listen to the sound of birds. The goal of this work is to break that intuition down, to trace it through neural activity, and to express it in equations grounded in the physics of the brain.

In ***Chapter 1***, explore what happens when a simple visual cue turns into a conscious experience. How does information move through the brain? What sets apart unconscious processing from awareness? To explore these questions, we study the neural dynamics underlying this transition—looking for patterns that signal a shift

from localized activity to widespread ignition. We approach this by simulating thousands of brains, each with a different architecture, to identify the conditions under which this transition emerges.

In ***Chapter 2***, we shift focus to the structure of the way the information is stored in the cortex. If the brain does not rely on fixed modules, how does it encode even the simplest visual content? To answer this, we develop machine learning tools that uncover how mental content is distributed across cortical space. What emerges is not a clean hierarchy, but a complex motif of overlapping and competing geometric influences.

Finally, in ***Chapter 3***, we explore the origin of these macro-scale influences shaping sensory representations. How does the cortex self-organize into a functional blueprint capable of supporting rich, flexible mental life? We examine the spatial competition and global constraints that give rise to cortical gradients, and argue that these mechanisms may provide a general scaffold for organizing any form of mental representation.

What follows is the path my intuitions led me to explore.

1

Distributed Dynamics Underlying Conscious Access

*The work described in this chapter led to the publication of the article **"A dynamic bifurcation mechanism explains cortex-wide neural correlates of conscious access"**, published in Cell Reports. Annex 1 presents the full paper.*

1.1 What Do We Mean by Consciousness?

> *The whole idea is based on a fundamental misconception that just because there is a noun 'consciousness' there is some 'thing' like magnetism or electricity or pressure or temperature, and that it's worth looking for correlates of that thing.*
>
> Aaron Sloman

If we want to understand what lies behind our intuition of consciousness, we first need to untangle the meaning of the word itself. And that turns out to be more difficult than it seems. Ask two people—even two scientists who study consciousness—and you're likely to get completely different answers. In English, the term can refer to self-awareness—"I think, therefore I am" (Descartes, 1641)—or to meta-cognition (Fleming and Dolan, 2012), as in "I know that this dissertation bores me, and I wish it were over already." It can also carry moral connotations, as in "acting with a clear conscience." (Sorabji, 2014) These meanings are clearly related, but they are distinct in a subtle way—and likely arise from different yet untangled neural mechanisms

In this work, I focus on just one of these meanings—though perhaps my scientific path will lead me to explore the others in time. To define it, we can return to William James and his notion of the "stream of consciousness." (James, 1890) At every moment,

the brain is bombarded with external sensory inputs as well as self-generated memories, emotions, and internal thoughts. All of this contributes to a continuous flow of information, and it would be overwhelming if all of it was analyzed extensively. Instead, only a subset is captured and directed into the stream of consciousness—the portion we experience and relate to.

Most of this information flowing through the brain is processed unconsciously — sometimes at very basic levels. But once a piece of information is captured by conscious access, it becomes available to a wider range of cognitive processes (Block, 1995). We can reflect on it, analyze it, draw conclusions. Take the well-known example of the gorilla walking among basketball players (Figure Figure 1.1) (Simons and Chabris, 1999). The visual information was present all along, entering the visual cortex. But once it is accessed consciously, we can perceive it as absurd, infer it was intentional, and recognize that the real experiment was about awareness, not ball-passing.



Figure 1.1: *Illustration of the classic "Invisible Gorilla" experiment (Simons & Chabris, 1999). In this task, participants are asked to count the number of passes made by players in white shirts. While focusing on this task, many fail to notice a person in a gorilla suit walking through the scene — a striking example of inattentional blindness.*

I focus on this narrow but rich sense of consciousness because it satisfies several key criteria: it describes a cognitive function with clear boundaries; it aligns with our intu-

itions; and most importantly, it is empirically testable. We can test whether something distinct happens in the brain when information becomes consciously accessed.

And something does happen (van Vugt et al., 2018; Sergent et al., 2021, 2005; Del Cul et al., 2007; Lamme, 2015). When a brief visual stimulus is presented—whether to a human, monkey, or rat—it triggers an automatic burst of activity in the early visual cortex. The activity then quickly dissipates; conscious perception makes little difference at this stage. In the associative cortex, however, the story changes. If the stimulus is consciously perceived, activity does not fade—it bifurcates. The network either returns to baseline or ignites into a strong, sustained activation. This transition has been observed at two complementary scales: in monkeys, with precise, intracranial recordings showing population-level bifurcation in prefrontal area 9/46d during report-based tasks (van Vugt et al., 2018) (Figure 1.2); and in humans, through EEG and MEG, revealing cortex-wide ignition patterns (Sergent et al., 2021; Del Cul et al., 2007). This all-or-none bifurcation dynamic (Sergent et al., 2021)— vanishing and igniting—has become a leading candidate for the neural signature of conscious access.

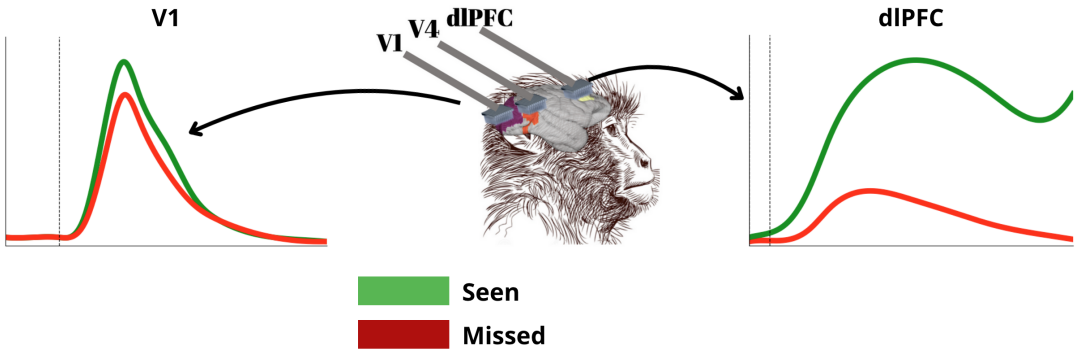

Figure 1.2: *Neural activity in visual and prefrontal cortex during conscious perception. Recordings from monkey visual cortex (V1) and dorsolateral prefrontal cortex (dlPFC) show distinct dynamics when stimuli are consciously perceived (green) compared to when they are missed (red). While both seen and missed trials elicit an early transient response in V1, only seen trials lead to a strong and sustained activation in dlPFC. Adapted from van Vugt et al. (2018).*

Why is this bifurcation such a compelling clue? Because the two branches tell very different stories about the fate of the same sensory input. When the neural wave fades, a weak and short-lasting trace of the stimulus can still be seen in prefrontal population activity. This may support some limited unconscious processing. When the network ignites, by contrast, the very populations that encode the stimulus content—its orientation, phoneme or face identity—enter a high-gain state. Their activity is selectively amplified, reverberates through associative and premotor areas far beyond the original

sensory maps, and can be encoded for several seconds after the stimulus is gone. This kind of widespread, lasting activation aligns well with what we intuitively expect from consciously perceived stimuli: that they leave a stronger trace, shape decisions, and become accessible to reflection, memory, and report.

If the bifurcation is indeed the neural correlate of conscious access, then we should be able to move beyond our intuitions and trace how physical processes in the brain give rise to this ignition—how physics produces conscious access.

This brings me to another reason I find this project so compelling. From the outset, I believed it will work, because I hold a physicalist view: that all of cognition arises from neuronal activity. But I'm also an agnostic, I acknowledge I could be wrong. Perhaps Descartes was right, and some non-physical substance intervenes (Descartes, 1984). Maybe, as we study the brain in increasing detail, we will predict exactly how sensory areas respond, how activity propagates—but at some point, a signal will diverge from the laws of physics. Perhaps it will happen in the pineal gland. Perhaps in the parietal cortex. I don't expect that outcome. But if intuitions about dualism are to be true, if there really is something beyond the physical, then it should leave a trace we could capture.

1.2 The Physics of Neurons

> *Everything should be made as simple as possible, but not simpler*
>
> Attributed to Albert Einstein

To derive conscious access from the laws of physics—just as life was explained through chemistry—we must first choose the right level of description. In our case, the challenge is to capture a specific bifurcation: a sudden, all-or-none, sustained transition in cortical dynamics associated with conscious access. While such bifurcations likely occur throughout the brain, reflecting other cognitive or physiological shifts, we focus on the one that aligns specifically with the neural correlates of conscious access. In biology, the key to understanding metabolism was to zoom in on the molecular scale; specifically, proteins and enzymes that orchestrate chemical cascades from cell division to sugar metabolism. This scientific field succeeded not because it reduced everything to one level, but because it began at the scale where the crucial dynamics emerged (Jacob and Monod, 1961). Likewise, to understand the physical emergence of conscious access, we must choose a level of analysis that captures the bifurcation—not too coarse, not too fine—and build from there.

Should we simulate the whole brain? Individual regions? Cortical columns, neurons, or even microtubules within them? I don't think that there's a single correct answer;

the appropriate level depends on the specific dynamics we aim to explain. But for our purposes, a natural place to begin is the neuron. Neurons resemble fluid-filled balloons floating in the extracellular fluid, with distinct ions inside and outside. This ionic imbalance naturally gives rise to a membrane potential—a voltage difference across the cell membrane—typically around −70 millivolts. Ions naturally diffuse across the membrane, working to balance the voltage between the inside and outside of the cell. In addition, membrane channels can open, permitting a strong, directional flow of specific ions that shifts the membrane potential upward or downward. When the membrane potential crosses a critical threshold, the neuron rapidly depolarizes, producing a sharp voltage spike known as an action potential. This electrical spike propagates along the axon and can activate other neurons downstream, creating the basis for communication and coordination across the brain.

Just as Newton transformed falling apples into equations (Newton, 1687), Hodgkin and Huxley turned spikes and ions into a formal mathematical description (Hodgkin and Huxley, 1952). Their model captured how ion flows, driven by the voltage difference between inside and outside the cell, evolve over time, and can lead to action potentials. Crucially, this means we can now describe neuronal behavior in precise mathematical terms and predict how a neuron would respond to different inputs. These equations showed that a single neuron can behave in a highly nonlinear way. When stimulated, it may remain silent, respond with a graded voltage shift, or suddenly fire an action potential. But this nonlinearity alone is not enough. In isolation, a single neuron quickly returns to baseline after spiking: it cannot sustain activity over seconds, nor can it flexibly encode or switch between distinct representations. The spike itself carries little semantic content. There is, as far as we understand, no mechanism by which an isolated neuron can generate a bifurcation that reflects specific information, or selectively amplify one representation over another.

To do that, we need to go beyond single neurons and connect them within a network, a structured population of neurons connected through synapses. When a neuron fires, the resulting action potential travels down its axon and triggers the release of neurotransmitters at the synapse. These chemical messengers bind to receptors on the postsynaptic cell, opening ion channels that shape the response of the receiving neuron. Theoretical neuroscience has extended the Hodgkin–Huxley framework to model not just the membrane potential of individual neurons, but also the dynamics of synaptic transmission (Katz and Miledi, 1967; Destexhe et al., 1994). In these models, each type of synaptic receptor—AMPA, NMDA, and GABA—is governed by the same differential equations, differing only in their time constants. AMPA and NMDA receptors are both excitatory, increasing the likelihood that the receiving neuron will fire, while GABA receptors are inhibitory, tending to suppress such activity. Their effects unfold over distinct timescales: AMPA channels typically stay open for 2 milliseconds, GABA for 5 milliseconds, and NMDA for 60 milliseconds. Notably, NMDA receptors are also

voltage-dependent, requiring some depolarization to fully activate. These differences in timing and sign play a crucial role in how neural networks maintain, amplify, or suppress patterns of activity.

1.3 Bifurcation in a Neural Population

Studying a single neuron is already a formidable task. To predict its behavior, we must track its membrane potential over time, along with the state of various ion channels on its surface. Even then, we are still relying on approximations. Now imagine scaling that up to thousands of neurons, each interacting with others through dense webs of connections. Simulating such a system is already challenging for a computer. For a human trying to understand it, the complexity becomes unmanageable. One of the most elegant approaches, in my view, was introduced by Wong and Wang in 2006 (Wong, 2006). They showed that a network of thousands of neurons—comprising both excitatory and inhibitory cells—can be reduced to just two variables: the fraction of NMDA open channels on the excitatory population's axon terminals, and the fraction of GABA open channels on those of the inhibitory population. We used this framework as our starting point, to investigate whether a bifurcation—like the one observed during conscious access—could emerge from these simplified yet biologically grounded dynamics.

The main advantage of this simplified model is that it reduces the network to a system governed by just two dynamical equations (Equation 1). These equations specify how the state of the system evolves over time—given its current configuration, we can precisely predict its trajectory. This analytical tractability gives us a rare kind of clarity: we can determine, with confidence, what the system can and cannot do.

$$\frac{ds_N}{dt} = -\frac{s_N}{\tau_N} + (1 - s_N)\gamma_N r_E \tag{1.1}$$

$$\frac{ds_G}{dt} = -\frac{s_G}{\tau_G} + (1 - s_G)\gamma_G r_I \tag{1.2}$$

$$I_E = G_{E,N} s_N - G_{E,G} s_G + I_{E,bg} \tag{1.3}$$

$$I_I = G_{I,N} s_N - G_{I,G} s_G + I_{I,bg} \tag{1.4}$$

$$r_E = \frac{a I_E^{total} - b}{1 - e^{-d(a I_E^{total} - b)}} \tag{1.5}$$

$$r_I = \left\{ \begin{array}{ll} \beta(I_I^{total} - I_{th}) & \text{for } I_I^{total} \geq I_{th} \\ 0, & \text{otherwise} \end{array} \right\} \tag{1.6}$$

The model is described by two coupled differential equations (Equation 1), which track the evolution of synaptic gating variables in an excitatory and an inhibitory population. The variables s_N and s_G represent the fraction of open NMDA and GABA receptor channels, respectively—serving as proxies for the activity of the excitatory and inhibitory populations. Each evolves with its own time constant, τ_N for NMDA and τ_G for GABA, and is modulated by the firing rate of the corresponding population. The parameters γ_N and γ_G control how quickly the input current influences the gating variables. The total synaptic current received by each population is a weighted sum of recurrent excitation and inhibition: I_E depends on self-excitation $G_{E,N}s_N$ and feedback inhibition $G_{E,G}s_G$, plus a constant background input $I_{E,bg}$; conversely, I_I reflects excitation from the excitatory population $G_{I,N}s_N$, self-inhibition $G_{I,G}s_G$ and background input $I_{I,bg}$. These equations capture how mutual interactions between the populations drive the system's dynamics—and, as we will see, under the right conditions they can produce bifurcations. a, b, d, β and I_{th} are constant modulated the impact of the currents on the population firing rates.

Because this system has only two variables—representing the fraction of open NMDA and GABA channels, each evolving between 0 and 1—we can visualize its full state space in two dimensions. This gives us a phase plane (Strogatz, 2015): a 2D plot where each point corresponds to a possible combination of excitatory and inhibitory activity. Since the system's equations (Equation 1) specify how s_N and s_G change over time, we can draw a vector field across this plane, showing the direction in which the system will evolve from any initial state (Figure 1.3). This allows us to visualize trajectories—how the system will move.

On this phase plane, we can plot the nullclines of the system: the set of points where the rate of change for each variable is zero ($ds/dt = 0ds/dt = 0ds/dt = 0$). In Figure 1.3, the red curve shows where the excitatory variable s_N is at equilibrium, and the blue curve does the same for the inhibitory variable s_G. Wherever these two curves intersect, both variables are simultaneously at equilibrium—meaning the system, if placed exactly at that point, will remain there unless perturbed.

In the current setup, the nullclines intersect at three points. The outer two are stable fixed points, or attractors: the vector field around them points inward, so the system naturally settles there. The middle one, however, is unstable: the arrows near it point away, so even a small perturbation will push the system toward one of the two stable states. In effect, it acts as a threshold between two qualitatively different behaviors. One attractor corresponds to low excitatory activity (small s_N), where the system remains near baseline. The other shows high s_N, meaning that the excitatory population enters a self-sustaining, strongly active state—driven by positive feedback within the circuit.

Of course, these dynamics depend on the model's parameters: the strength of recurrent excitation, the balance of inhibition, and the time constants, among others.

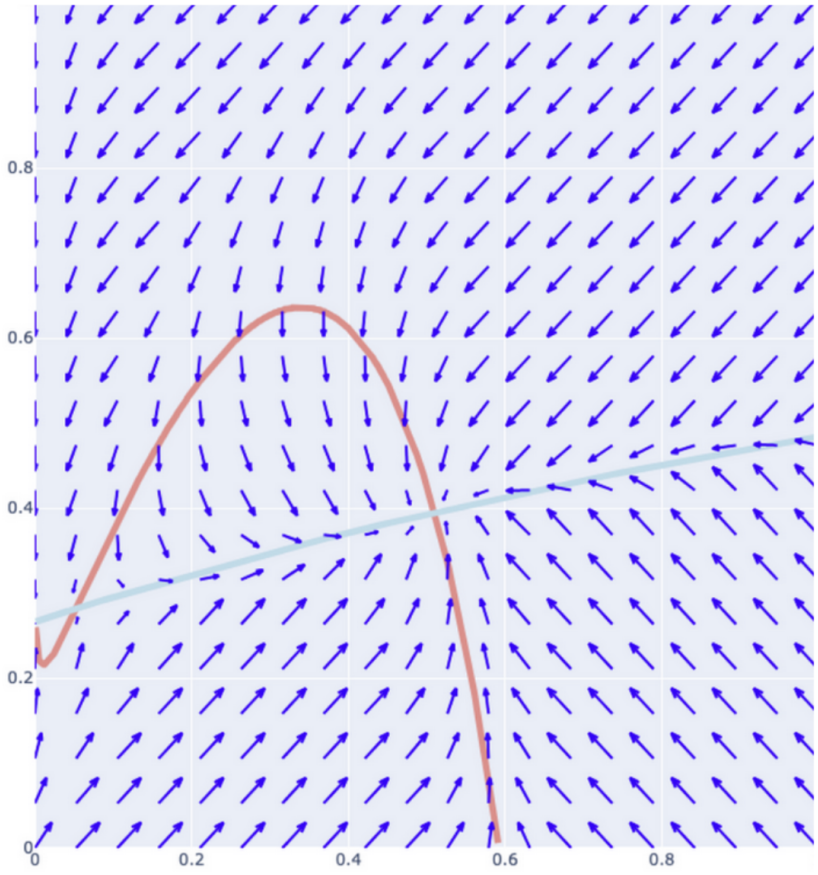

Figure 1.3: *Phase portrait of a two-population system. The blue curve represents the nullcline of the excitatory population, while the red curve represents the nullcline of the inhibitory population. Their intersections correspond to fixed points of the system. Arrows indicate the direction of trajectories in the phase space. Axes denote the normalized activity of the excitatory (x-axis) and inhibitory (y-axis) populations.*

Increasing excitation shifts the red nullcline upward, potentially eliminating the resting state and making high activity the only stable solution. Weakening excitation has the opposite effect, suppressing the high-activity attractor. This means the system's qualitative behavior—whether it supports a bifurcation—can be controlled by a few biophysically meaningful parameters. What this analysis provides is not just a simulation, but an understanding. By visualizing how the system behaves across all possible states, we gain deep insight into the conditions that allow for sudden, sustained shifts in neural population activity—the very kind of bifurcation that may underlie conscious access.

To go further, we can model what happens when the system receives a brief external stimulus. In practice, this amounts to increasing the excitatory background

input $I_{E,bg}$ in Equation 1. Analytically, we can track the consequence on the phase plane (Figure 1.4): this input transiently pushes the red nullcline (the excitatory one) upward. If the stimulus is strong enough, the resting state disappears altogether—leaving the system with only one stable attractor. As a result, even a network that starts near rest will be forced to move toward the high-activity state. In simple terms, injecting current leads to increased firing. But the value of this model lies in how it explains why and under what conditions that increase becomes self-sustaining.

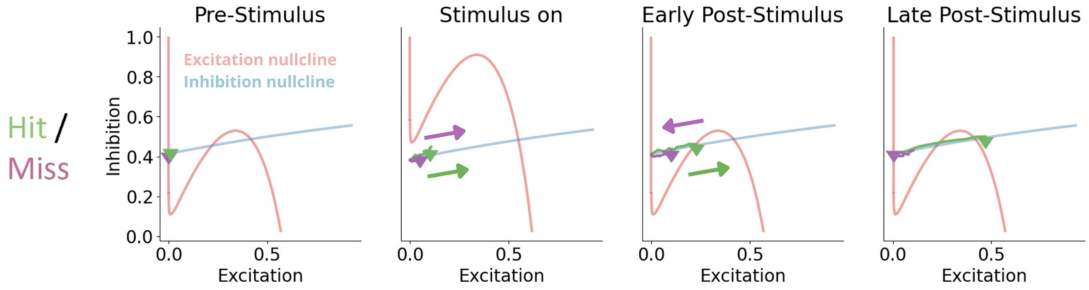

Figure 1.4: *Phase portraits of a simplified excitatory–inhibitory network across trial stages. Each panel shows the nullclines of the excitatory (red) and inhibitory (blue) populations, with trajectories for individual hit (green) and miss (purple) trials.* ***Pre-stimulus:*** *activity rests near the low-activity steady state.* ***Stimulus on:*** *the excitatory nullcline shifts upward, reducing the number of intersections with the inhibitory nullcline and allowing trajectories to diverge.* ***Early post-stimulus:*** *in hit trials, activity enters the basin of attraction of the high-activity steady state, while in miss trials it remains near the low-activity state.* ***Late post-stimulus:*** *after stimulus removal, nullclines return to baseline, with trajectories progressing toward their respective attractors.*

Once the stimulus is removed, $I_{E,bg}$ returns to its baseline value, and the red nullcline returns to its original position. At that point, the system finds itself somewhere in the original phase plane. Whether activity persists or fades depends on where it lands: if the trajectory has crossed the unstable fixed point, the system is now in the basin of attraction of the excited state and will continue converging toward high activity. If not, it will fall back toward rest. This creates a natural all-or-none dynamic—either the stimulus was strong enough to tip the system over the threshold, or it wasn't.

This simple model captures the core features of bifurcation, and it does so in a way that is analytically transparent. We call this a dynamic bifurcation, because unlike traditional bifurcations—where a fixed parameter change alters the number or stability of equilibrium points—here, the phase plane itself evolves over time. As the input rises and falls, the system's structure temporarily shifts, and the outcome depends on whether the trajectory crosses a moving threshold. The stronger the input, the larger the shift, and the more likely the system is to transition to the excited state(Figure 1.5.

In this way, a continuous external variable (stimulus strength) produces a discontinuous internal effect—a hallmark of conscious access.

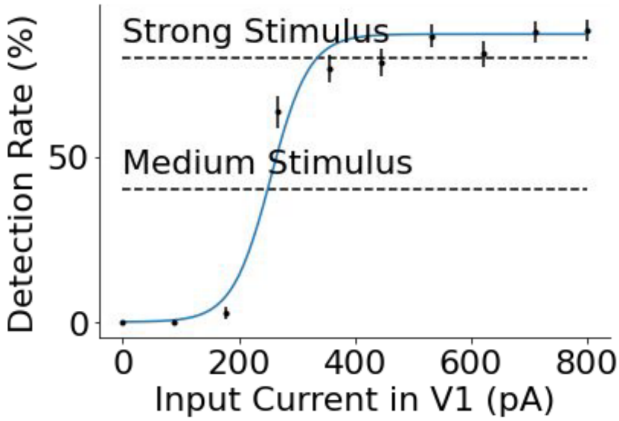

Figure 1.5: *Sigmoidal relationship between stimulus intensity and detection probability. Detection rate in the large-scale cortical model increases nonlinearly as a function of input current delivered to V1. Medium and strong stimulus regimes correspond to approximately 50% and 80% detection rates, respectively. Data points represent detection rates across trials, with error bars indicating variability.*

This simple system already captures the essential dynamics of hits and misses in response to faint stimuli. But its behavior can be pushed further. By increasing the system's "vigilance"—modeled as a sustained increase in the background excitatory input $I_{E,bg}$-we make the network more responsive. Fainter stimuli now have a better chance of pushing the system past the bifurcation threshold. However, this comes at a cost: the basin of attraction for the resting state shrinks. As a result, the system becomes more likely to spontaneously escape the rest state and transition into the excited state—even in the absence of any external input.

These dynamics offer a natural explanation for false alarms observed in perceptual experiments, where participants occasionally report seeing a stimulus that was never presented. Pushed further, the same mechanism may provide an analytical foothold for understanding hallucinations more broadly. When spontaneous transitions into self-sustaining activity become likely—either due to heightened internal drive or weakened inhibition—the system may generate vivid perceptual states that are internally triggered but externally ungrounded. These ideas are developed more fully in a forthcoming paper currently in preparation.

These results show how even a single population of neurons can exhibit bifurcation dynamics. But this is only a small part of the picture. If dynamic bifurcation truly underlies conscious access in the brain, we must go further: Which neuronal populations are responsible for it? How distributed is this process across the cortex? And how is the underlying dynamic shaped by external stimuli and inter-regional connectivity?

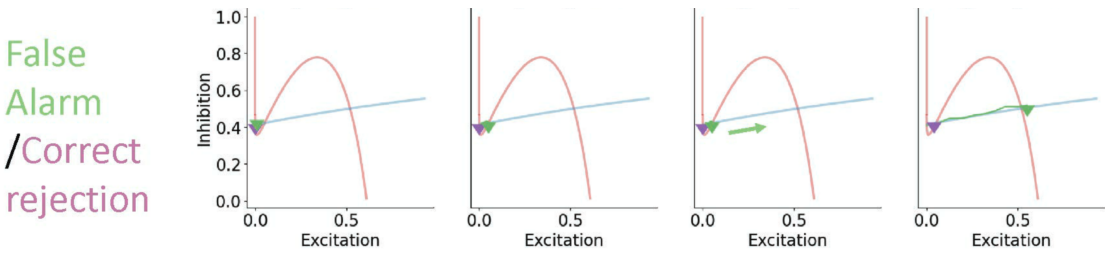

Figure 1.6: *Phase portraits of the excitatory–inhibitory network under vigilance without stimulus. The excitatory (red) and inhibitory (blue) nullclines are shown across trial stages. A vigilance signal slightly shifts the excitatory nullcline upward, mimicking a subcortical input. Example trajectories illustrate two possible outcomes: in* ***false alarm trials*** *(green), activity crosses into the basin of attraction of the high-activity state despite the absence of a stimulus; in* ***correct rejection trials*** *(purple), activity remains in the basin of attraction of the low-activity state and returns to baseline.*

To answer these questions, we need to move beyond minimal circuits and investigate how these mechanisms unfold at the scale of cortical networks.

1.4 Bifurcation in the Cortex

> *Equations are not my language. I'm not saying that this is not an important way to look at life. Just that it isn't my way, and I'm not equipped to answer questions on it.*
>
> Richard Dawkins

As we saw in the previous section, bifurcation does not emerge at the level of individual neurons, but it can arise at the level of a neural population. If dynamic bifurcation truly underlies conscious access, however, we must understand how it manifests at the scale of the cortex—a far more complex problem. At this scale, the brain is composed of many interacting regions, each containing its own population dynamics. Once we move beyond a single pair of excitatory and inhibitory populations, we can no longer rely on two-dimensional phase plane analysis. To tackle this challenge, we must shift from analytical tools to computational modeling. Like Dawkins before me, I find myself at the edge of my mathematical fluency.

Our approach involves simulating networks of interconnected neuronal populations, each governed by the same dynamical equations described earlier. We then explore how bifurcation dynamics emerge in different parts of the cortex, and under what conditions. This differs fundamentally from machine learning approaches, where

artificial networks are trained to perform specific tasks (LeCun et al., 2015). Instead, our methodology resembles that of climate science, where computer simulations of wind, pressure, and temperature—grounded in physical law—are used to forecast weather patterns (Washington and Parkinson, 2005). In the same spirit, we aim to simulate the evolution of cortical activity over time by reproducing the physical dynamics of neural populations distributed across the brain. Specifically, we aim for the simulated network to reproduce key population-level signatures observed in the empirical study of van vugt et al. (2018) (van Vugt et al., 2018), where a brief visual stimulus elicits a transient burst of activity in early visual cortex (V1), but can trigger a sustained, all-or-none activation in prefrontal area 9/46d. This divergence in temporal dynamics—transience in sensory regions, bifurcation in associative cortex—serves as a functional benchmark for our model. It allows us to test whether the equations and parameters governing local populations, when embedded in realistic cortical architecture, are sufficient to generate the large-scale dynamical transitions associated with conscious access.

To build our simulated cortex, we drew on the foundational work of researchers from Xiao-Jing Wang's lab (Froudist-Walsh et al., 2021; Murray et al., 2014; Joglekar et al., 2018; Mejías and Wang, 2022), who developed large-scale models of the macaque brain. In our model, the cortex is represented as a set of interconnected regions, each containing local receptive fields that compete through a winner-take-all mechanism and interact via long-range projections. Specifically, each region includes two local excitatory populations, each corresponding to a different sensory representation. These populations project locally, activating both NMDA and AMPA receptors, and they compete through a shared inhibitory population. This inhibitory loop ensures that strong activation of one representation suppresses the other. While this setup is a simplification—it clearly doesn't make sense for every cortical region to host all possible sensory representations—it provides an operational framework for exploring the dynamics of competition between neural representations selection at the cortical scale.

To simulate inter-regional interactions, we included long-range excitatory connections between cortical areas, constrained by anatomical data. These projections were informed by the work of Markov et al. (2013) (Markov et al., 2014), who injected retrograde tracers into 40 different cortical regions. These tracers travel backward along axons, labeling the neurons throughout the cortex that send projections to the injection site. This dataset provides a rich, empirically grounded map of directed and weighted connectivity between regions, which we used to define the structural backbone of our model.

We started with the assumption that anatomical connectivity alone—the raw connection strengths between regions—would not be enough to reproduce the bifurcation dynamics observed in conscious access. The cortex is not just a web of connected

regions; it is hierarchically organized into sensory processing streams, where information flows from primary sensory areas toward higher-order associative regions. These pathways are not symmetric. Connections from lower to higher regions are called feedforward, while those going in the opposite direction are feedback. Crucially, the same Markov et al. (2013) dataset provides not only connection weights but also the laminar distribution of afferent neurons, dividing each projection across superficial, middle, and deep cortical layers (Figure 1.7). Along known sensory hierarchies, feedforward connections typically originate in the upper layers of lower-level areas, while feedback connections tend to arise from the deeper layers of higher-level areas (Felleman and Van Essen, 1991; Markov et al., 2014). This laminar pattern allowed Markov and colleagues to assign each connection a feedforwardness score—quantified as the fraction of labeled neurons located in the supragranular (upper) layers.

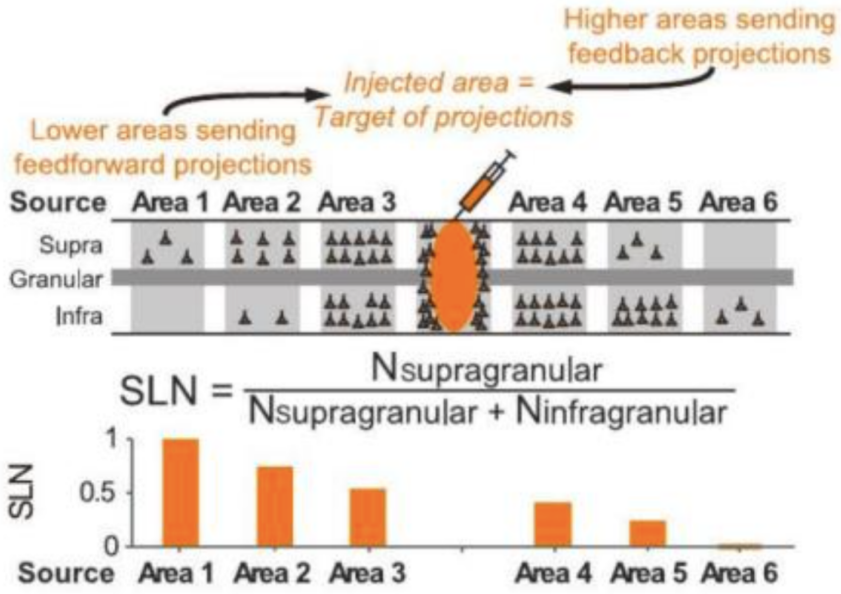

Figure 1.7: *Laminar origin of feedforward and feedback projections in the macaque cortex. An injection in a target area (orange) labels projecting neurons in source areas. Projections originating predominantly from supragranular layers (upper layers) correspond to* ***feedforward*** *connections, whereas those originating from infragranular layers (deeper layers) correspond to* ***feedback*** *connections. The fraction of supragranular neurons, quantified as the supragranular labeled neuron (SLN) index, provides a continuous measure of hierarchical distance between cortical areas. Adapted from Markov et al. (2014).*

They found that no connection is purely feedforward or purely feedback; instead, each inter-areal pathway contains a mix of upper-layer and deeper-layer projections, placing it at a specific point along a graded continuum of directional influence. This balance is quantified by the supragranular layer neuron (SLN) value, which is low

for well-established feedback connections and high for classical feedforward ones. Similarly, lateral connections—between areas at similar hierarchical levels—tend to have intermediate values around 0.5, reflecting a more balanced laminar distribution. This gradient of directional influence became a key feature of our model. Yet, while the dataset from Markov and colleagues provides detailed information about inter-areal connectivity and laminar origin, it does not tell us how these connections are implemented at the synaptic level. In particular, we lack data on which neuronal populations—excitatory or inhibitory—these long-range projections target, and through which types of excitatory receptors, such as AMPA or NMDA, their effects are mediated. These experimental gaps leave crucial model parameters unconstrained, allowing us to explore a plausible range of configurations within the unknown parameter space.

To model hierarchical processing without overfitting—and to keep the model as simple and interpretable as possible—we introduced a minimal set of assumptions. We treated all projections originating from upper layers (regardless of source region) as sharing the same two properties: (1) a fixed AMPA/NMDA receptor ratio, and (2) a fixed targeting ratio between excitatory and inhibitory populations in the receiving area. We applied the same logic to deep-layer projections: all feedback-like projections shared a common AMPA/NMDA receptor ratio and excitatory/inhibitory targeting fraction. This gives us just four parameters to govern inter-areal synaptic dynamics across the cortex: two for projections originating in the upper layers, and two for those from the deeper layers. For any given connection, the exact AMPA/NMDA receptor ratio and excitatory/inhibitory ratio were computed as weighted averages based on the connection's feedforwardness value, derived from the SLN (supragranular labeled neurons) matrix. This approach allowed us to encode directionality and synaptic specificity in a principled, low-dimensional way, tightly grounded in empirical data.

By systematically exploring this four-dimensional parameter space, we identified three distinct dynamical regimes that emerged when a brief stimulus was applied to one of the receptive fields in V1, the primary visual cortex. In the vast majority of configurations, the network responded to the stimulus with a transient burst of activity, then quickly returned to baseline across all regions—failing to sustain or propagate the information. At the opposite extreme, certain parameter combinations led to runaway excitation: the entire cortex, including V1, entered a high-activity state resembling pathological overactivation, akin to a seizure. But in a narrow region of the parameter space, we observed something qualitatively different—and strikingly consistent with empirical data from van vugt and colleagues. In this regime, V1 responded with a brief, stereotyped activation that reliably returned to baseline, regardless of trial. Crucially, in the prefrontal area 9/46d, we observed a bifurcation: across identical stimulus presentations, the same region either returned to rest or transitioned into a high, sustained activity state—mirroring the all-or-none ignition pattern associated with conscious perception. This was the only regime that reproduced both the local fading

response in the sensory cortex and the bistable ignition dynamics in the prefrontal cortex—a key signature of bifurcation(Figure 1.8).

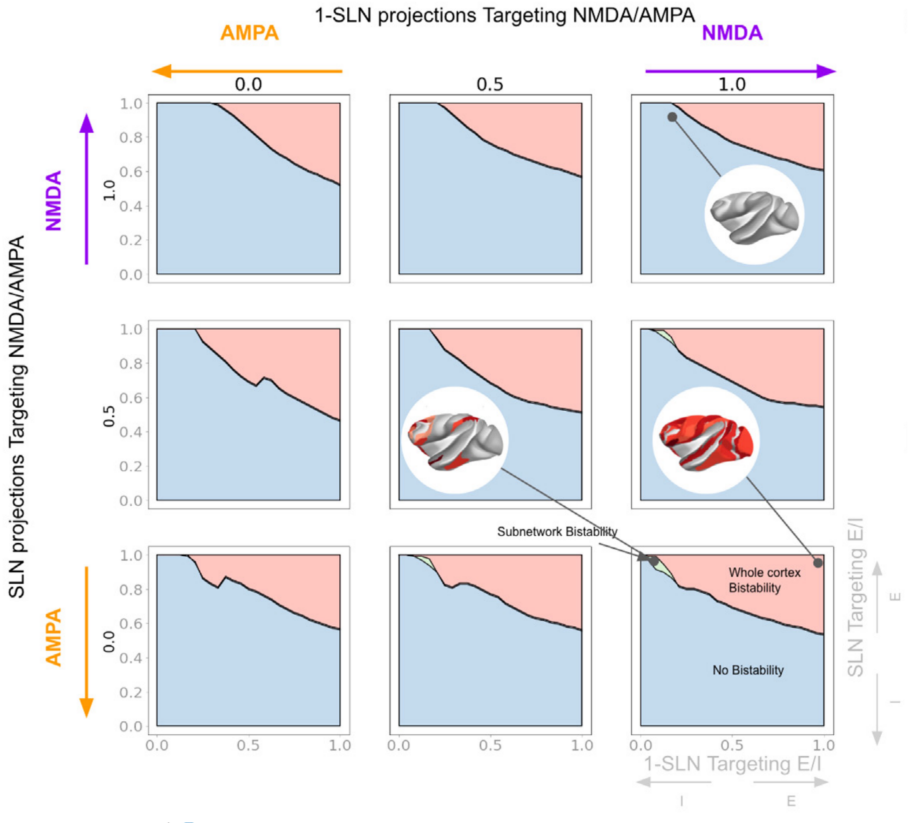

Figure 1.8: *Parameter search of inter-areal pathways. Each panel shows a two-dimensional parameter search over glutamatergic receptor contributions (AMPA vs. NMDA) and cellular targeting (excitatory vs. inhibitory) for inter-areal connections. Rows correspond to different fractions of SLN-mediated projections targeting NMDA vs. AMPA, while columns correspond to 1–SLN projections. The models exhibit three distinct dynamic regimes:* ***subnetwork bistability*** *(green),* ***whole-cortex bistability*** *(red), and* ***no bistability*** *(blue). Example parameter sets are mapped onto this space (insets) to illustrate the corresponding network-level activity patterns.*

1.5 Introducing Atoum: A Synthetic Cortex for Modeling Awareness

We named the network Atoum (Klatzmann et al., 2025), in reference to the first Egyptian god who became conscious, as it occupies the narrow region of parameter space where awareness-like bifurcation dynamics emerge. Within this four-dimensional space,

Atoum exhibited the precise transitions we set out to capture. Although we tuned the system based only on the behavior of two regions—V1, representing early visual processing, and area 9/46d in the prefrontal cortex—we observed structured dynamics across the entire cortical network. A brief (50 ms), faint stimulus to one of V1's receptive fields triggered a cascade of activation along the dorsal and ventral visual streams (Felleman and Van Essen, 1991). The activity propagated sequentially through the visual hierarchy: V1, V2, V4 and finally LIP. This wave then triggers the ignition of a distributed associative network composed of 17 of the highest regions in the cortical hierarchy(Figure 1.9). These areas responded as a unified system, collectively exhibiting a bifurcation dynamic: in some trials, activity quickly faded; in others, it crossed a threshold, ignited, and stabilized, maintaining elevated activity for several seconds—effectively holding the stimulus in working memory. This suggests that large-scale functional organization can emerge purely from the architecture of inter-areal connectivity and the tuning of receptor-level synaptic parameters. Spatially structured cortical dynamics can arise from biophysical and anatomical constraints alone, without any need to hard-code spatial maps.

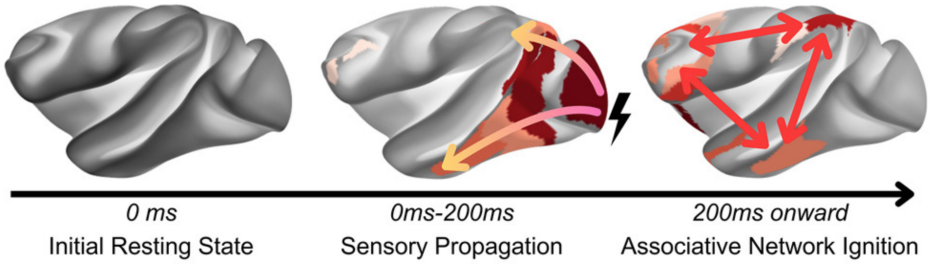

Figure 1.9: *Cortex-wide ignition during conscious perception. Schematic of the temporal evolution of large-scale cortical dynamics following a sensory input. At **0 ms**, the system is in a resting state. Between **0–200 ms**, sensory activity propagates from early sensory regions along feedforward pathways. Beyond **200 ms**, recurrent interactions among associative and prefrontal regions lead to widespread, sustained activation — a hallmark of cortical ignition associated with conscious access.*

Perhaps most strikingly, the ignition of associative regions was not engineered into the model—it emerged spontaneously from the interaction between local nonlinear dynamics and large-scale cortical connectivity. This closely mirrors the central prediction of Global Neuronal Workspace (GNW) theory (Mashour et al., 2020; Boly et al., 2017; Dehaene et al., 2011), which posits that a distributed network of associative neurons—termed the workspace—serves to broadcast the content of local neuronal processors across the brain, enabling global accessibility and conscious access.

Another influential, though more controversial (Doerig et al., 2019; IIT-Concerned et al., 2023), theory of consciousness—IIT, the Integrated Information Theory—proposes

that a physical system gives rise to conscious experience when it is maximally irreducible (Tononi, 2004; Oizumi et al., 2014; Albantakis et al., 2023). That is, when its internal causal interactions cannot be decomposed into independent parts and the system can occupy a rich repertoire of distinct states. A canonical example of such a system is a majority network (Oizumi et al., 2014), in which each neuron updates its state based on the dominant input from the population. Intriguingly, the subset of associative regions that exhibit bifurcation dynamics in our model form exactly this kind of majority network. While we did not explicitly compute Φ (the IIT measure of irreducibility), it is plausible that this subnetwork constitutes the system's maximally integrated core. This observation raises the provocative possibility that GNW and IIT—often treated as competing theories—may instead describe complementary aspects of conscious processing. GNW may capture the temporal dynamics that enable access to a percept, while IIT characterizes the structural features of that percept once it is instantiated in experience. Moreover, while IIT has traditionally associated high-Φ complexes with posterior cortical regions (Boly et al., 2017)—where local circuitry may be strongly self-influencing—this focus may miss the broader picture. Our findings suggest that the largest integrated complex might span a distributed network, including prefrontal and associative regions. In this light, conscious access and conscious experience may reflect the same physical substrate, viewed through the lenses of dynamic ignition and structural integration, respectively.

Now that we have a functioning brain model running on a computer, we can begin to treat it as a real experimental subject. We can present stimuli, record from any area, perturb the system in controlled ways, and observe its internal dynamics with complete transparency. Every variable, from firing rate, synaptic current to receptor state, is accessible at every time point and across the entire network. Of course, the model is a simplification: many biological details are abstracted away. But even in this simplified form, it already displays rich and interpretable dynamics. This has allowed us to simulate a range of classic consciousness experiments, including priming, backward masking, and even cross-modal integration—successfully replicating qualitative patterns reported in the empirical literature. We have also begun to probe the model in more speculative directions, simulating lesions by silencing specific neuronal populations (mimicking trauma or stroke), and the injection ketamine, a known NMDA receptor antagonist (Zorumski et al., 2016), to explore the network's behavior under pharmacological disruption. While these extensions are still preliminary, they demonstrate the power of the model as a scientific platform—not just for reproducing known effects, but for generating mechanistic insights into how cortical dynamics unfold under varying conditions. It opens the door to a new kind of theoretical neuroscience: one in which we can run controlled experiments on a synthetic brain whose equations we fully understand.

The long-term hope is that such models could eventually serve as testbeds for

clinical interventions (Jirsa et al., 2023). By simulating the effects of specific drugs, lesions, or surgical modifications, we may one day predict how a given manipulation could alter large-scale brain dynamics—and, by extension, cognitive function. While we're far from clinical application, the foundation is now in place: a mechanistically interpretable, physiologically grounded cortical model that responds realistically to stimulation and perturbation. With it, we can begin to ask not just how the brain works, but how we might safely and precisely intervene when it doesn't.

1.6 Predicting the Brain

To build Atoum, we integrated extensive empirical data to constrain the model biologically. Yet in key areas—particularly the synaptic properties of long-range connections—direct experimental measurements were lacking. This forced us to explore the parameter space to identify configurations that produced ignition-like dynamics. If Atoum truly captures the ignition dynamics of the macaque brain, then the parameter values it converged on likely reflect underlying biological reality—offering a unique opportunity to learn about cortical biology through simulation.

Strikingly, the subset of the parameter space that supports bifurcation is extremely narrow, placing tight constraints on the underlying synaptic architecture (Figure 1.10). For connections originating in superficial layers—the archetypal feedforward projections—we found they must be almost exclusively mediated by AMPA receptors and must target excitatory populations in their destination areas. This is not a surprising outcome: for a stimulus to be consciously perceived, the neural signal it generates must propagate rapidly along the visual streams to ignite associative networks. If these feedforward projections are too slow (i.e., NMDA-dominant) or target inhibitory populations, the signal either dissipates or fails to reach ignition threshold. Conversely, for connections arising from deep cortical layers—the classical feedback pathways—we found that they must target both excitatory and inhibitory populations, with a net effect that is modulatory rather than purely excitatory. These connections must include a substantial NMDA component, making them slower and more integrative than their feedforward counterparts. This temporal profile is critical: if feedback inhibition were too fast (i.e., AMPA-dominant), it would suppress activity before ignition could stabilize. Instead, the feedback architecture supports focal excitation of task-relevant circuits while suppressing competing or lower-priority representations, helping to stabilize the percept that has gained access. Importantly, all inter-areal connections in the model are a mix of superficial and deep layer projections. Thus, every pathway combines feedforward and feedback properties to varying degrees. Classic feedforward connections—those between adjacent hierarchical levels—are dominated by superficial inputs, propagating information rapidly and robustly. Classic feedback pathways, by

contrast, are deep-layer dominant and act primarily as modulatory influences on earlier processing stages.

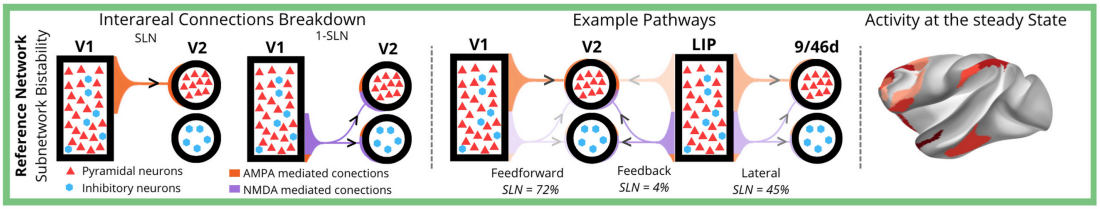

Figure 1.10: *Interareal connection architecture and example pathways.* ***Left:*** *schematic depiction of pure superficial-layer (SLN) and pure deep-layer (1–SLN) projections, indicating both target cell types (red = pyramidal/excitatory neurons, blue = inhibitory neurons) and glutamate receptor mediation (orange = AMPA, purple = NMDA).* ***Middle:*** *example pathways: a feedforward connection (V1 → V2, SLN = 72%), a feedback connection (LIP → V2, SLN = 4%), and a lateral connection (LIP → 9/46d, SLN = 45%). For each case, the SLN fraction determines the balance of superficial vs. deep projections and their receptor-mediated components.* ***Right:*** *cortical activity during an ignition for this parameter set, showing large-scale ignition.*

These synaptic constraints on inter-areal connections yield a novel and testable anatomical prediction. In the model, areas lower in the cortical hierarchy—such as primary sensory regions—receive a disproportionate amount of feedback input. To support the slow, modulatory influence of these feedback projections, these regions require a higher density of NMDA receptors. Conversely, associative regions higher in the hierarchy serve as major sources of feedforward output. To enable rapid signal propagation, they rely more heavily on AMPA-mediated transmission. This leads the model to predict a macro-scale gradient in the AMPA/NMDA receptor ratio: a high NMDA-to-AMPA ratio in the sensory cortex, gradually decreasing along the hierarchy toward associative areas(Figure 1.11. This prediction is counterintuitive. One might expect sensory areas to be dominated by fast AMPA transmission to process rapid sensory input, and associative areas to depend more on NMDA receptors to support slower, integrative cognitive functions. Yet the model suggests the opposite, based on circuit-level constraints necessary for ignition. With the help of our collaborators, we were able to test this prediction in macaque brain tissue using cortex-wide in vitro receptor autoradiography. Remarkably, the data confirmed the model's forecast: the NMDA/AMPA receptor ratio does follow a hierarchical gradient, aligning closely with the model's output.

This result is deeply exciting to me, not just because it validates the model, but because it represents a biological prediction derived purely from mathematical analysis of cortical dynamics. It shows that computational modeling can offer mechanistic insights into the brain's biochemistry, and that abstract dynamical principles can reveal hidden structure in real neural systems.

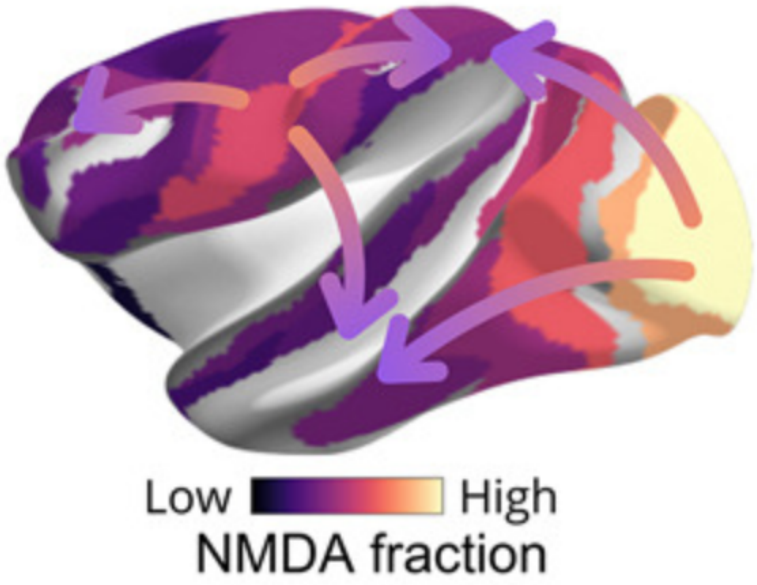

Figure 1.11: *Cortical distribution of NMDA receptor contributions. Surface map of the cortex showing the local fraction of NMDA-mediated excitation (color scale from low = dark purple to high = yellow). Purple arrows represent inter-areal pathways, highlighting that the fraction of NMDA contribution varies systematically across regions and influences large-scale dynamics.*

Two aspects of our results continue to puzzle me. First, our model revealed a macroscale gradient of receptor density—a striking pattern that echoes many known large-scale gradients of structural properties across the cortical surface. What makes this especially surprising is that the geometry of the cortex was never explicitly included in the model. The cortex was represented simply as a collection of areas connected by long-range projections, with no spatial information injected. And yet, the cortical geometry reemerges spontaneously in the solution, underscoring its fundamental role.

Second, the structure of mental representations caught my attention. In our model, every cortical area hosts a population representing each receptive field. This duplication—though necessary to reduce complexity and clarify the bifurcation structure—feels like an oversimplification. While the model offers insights into the dynamics of conscious access, it falls short of capturing the neural correlates of the conscious percept itself. Recent findings point to the distributed nature of mental representations, extending from lower sensory regions to higher associative areas—precisely those we identified as the bifurcation network. The degree to which these representations are distributed, how this distribution is constrained by cortical geometry, and how it shapes neural dynamics. These are questions I find impossible to ignore.

2

Spatial Influences on Cortical Sensory Representation

*The work described in this chapter led to the manuscript **"Spatial influences of the default mode and sensory networks behind distributed visual representations"**, which is currently under revision. The full preprint is presented in Annex 2.*

In Chapter 1, we examined how conscious access can emerge from large-scale cortical dynamics—how a percept, when unconsciously processed, remains confined to low-level sensory regions, but when captured by the associative network, becomes broadly broadcasted across the cortex. We saw how bifurcation dynamics enable this all-or-none transition, and how such ignition patterns may underlie the global availability characteristic of conscious perception. But there is a hidden assumption in our model that now demands scrutiny. In our current simulations, the "representation" of sensory content is radically simplified: every cortical area contains the same set of receptive fields, and these fields are structurally identical. The only functional difference between areas lies in their connectivity—not in what they represent. This uniformity, while useful for isolating ignition dynamics, is biologically implausible. It flattens the rich diversity of cortical selectivity into a homogeneous canvas. And if we want to understand how conscious access operates in a real brain—how specific contents are selected, integrated, and maintained—we must go beyond the ignition dynamics and examine the structure of mental representation itself.

How are representations distributed across the cortex? Are they localized, fragmented, or continuous? And how does this spatial organization shape what gets accessed—and why?

2.1 The Cortical Distribution of Mental Representation

The idea that the brain is composed of specialized parts—each performing a discrete cognitive function—has long shaped neuroscience (Brodmann, 2005; Glasser et al., 2016; Rakic et al., 2009). This "fractal" model of functional specialization finds support in anatomical mapping, lesion studies, and topographic representations from V1 to the fusiform face area. Yet as discussed in the Introduction, this framework breaks down in associative cortex, where boundaries blur, and representations become distributed, overlapping, and dynamic. At the population level, neural codes are shaped less by isolated modules than by shared, flexible ensembles that cut across anatomical divisions. This challenges the classical view of localization and calls for a new approach—one that can capture the graded, spatially extended nature of cortical representation. In the rest of this chapter, we explore how that organization might be systematically mapped.

A clear example of how mental representations are distributed across the cortex comes from retinotopic selectivity, the encoding of an object's location in the visual field relative to where we are looking (Wandell et al., 2007). Early work showed that neurons in primary visual cortex (V1) have preferred receptive fields: they respond most strongly when a stimulus appears in a particular part of space. These receptive fields are arranged topographically, so that nearby neurons tend to encode nearby locations in the visual field, and together form a map that encodes the full visual space. Because the entire visual field is laid out in this orderly way within a single cortical area, V1 became one of the classic examples supporting the fractal model of cortical organization. As imaging techniques improved, the picture began to expand: neighboring visual areas, such as V2, and V3 (Wandell et al., 2007), each contained their own retinotopic map. What had seemed unique to V1 turned out to be repeated, creating a layered representation of the visual field. Exploration then moved beyond early visual cortex, revealing retinotopic maps in higher-order regions along the dorsal stream, and even in associative areas such as the lateral intraparietal area (LIP) and the frontal eye fields (FEF) (Mackey et al., 2017). Today, with high-resolution fMRI and population receptive field (pRF) mapping, more than twenty distinct retinotopic maps have been identified across the brain, including in parietal and prefrontal regions (Benson et al., 2018; Himmelberg et al., 2021). This widespread distribution challenges the traditional view that retinotopy belongs only to early visual cortex. Instead, it suggests that spatial location is encoded in many parts of the brain, even in regions long thought to be dedicated to abstract or cognitive processing. Roughly eighteen percent of the cortical surface shows some degree of retinotopic organization (Benson et al., 2018). Yet these areas are far from representing spatial location alone, and it would be unreasonable to think that such a large portion of the cortex is devoted solely to retinotopy. Rather, in each of these regions, only a fraction of the local neural activity is dedicated to encoding visual field position. This can be quantified using a measure

we call retinotopic selectivity: the proportion of signal variance in each cortical vertex that can be explained by the spatial position of the stimulus (Dumoulin and Wandell, 2008). If this number is 100%, the vertex responds solely based on location. If it is 0%, retinotopy plays no explanatory role at all. This measure provides a continuous, graded description of functional specialization: not "does this region encode space?" but "how much of its activity is devoted to spatial encoding?" And across the cortex, the answer is nuanced. Even in regions with strong retinotopic organization, a large share of the activity cannot be explained by spatial information alone. Similarly, higher associative areas, where retinotopic selectivity is low, still carry a measurable trace of spatial information, showing that these regions integrate many kinds of signals, including sensory inputs. This suggests that every cortical area contributes to multiple functions, and every function draws on distributed, overlapping representations.

It's important to emphasize, however, that retinotopic selectivity simply measures how predictable a region's signal is based on visual input to the retina. This does not imply that the region is directly representing spatial location, or that this information is being used in that form by the brain. A region may be primarily involved in computing something else—like gaze direction or attentional focus—and still exhibit a statistical correlation with retinotopic input. To illustrate this point, consider a seemingly absurd but real example: given a high-resolution image of a person's face, it is possible to predict their blood type slightly better than chance. This doesn't mean the image "contains" blood type information in any functional sense, nor that the organism extracts or uses it. Yet, the very fact that retinotopic information is decodable from distributed cortical regions indicates that it is, in principle, accessible throughout the network. This accessibility suggests a coding architecture that is highly distributed and functionally entangled. Even when regions are not explicitly "representing" space, their dynamics remain constrained—at least in part—by retinotopic input. This reinforces the view that cortical representations are not modular containers, but dynamic patterns shaped by shared constraints across the cortical sheet.

If retinotopic selectivity offers a detailed, vertex-level map of how visual information is represented across the cortex, then in theory, we could build a model that mirrors it exactly—assigning one computational unit to each cortical vertex, around 30,000 per hemisphere. This would give us a one-to-one anatomical match between brain and model. But such a literal reproduction, while exhaustive, is conceptually limited. It's like trying to navigate a city using a satellite image: you see every building, shadow, and intersection, but none of the information that actually helps you move—no road types, no directions, no traffic flow. What you need instead is a GPS map: a simplified representation that preserves only the relevant structure, clarifies relationships, and reveals how the system operates. Likewise, a 1:1 cortical model may capture where information is encoded, but not how those patterns arise, how they interact, or what dynamics sustain them. It gives us a static picture, not a mechanistic account. Without

modeling the forces that shape cortical organization—competitive dynamics, spatial constraints, gradients in connectivity—we remain at the level of description, unable to extract the principles that govern functional structure. To understand how representations are distributed, accessed, and integrated, we need to move beyond reproduction and seek compression: models that reveal the logic beneath the surface.

So, can we move beyond a descriptive account of the retinotopic selectivity map? Can we reduce its complexity, uncover the logic behind it, and identify the forces shaping its structure? The highest retinotopic selectivity is found in early visual cortex, with V1 at the top of the list. This is no surprise, as it receives direct visual input from the lateral geniculate nucleus (LGN) (Callaway, 2005). Moving outward, mid-level sensory areas such as V4, V6, and V7 still show strong selectivity, but values tend to drop further in higher-order associative regions (Benson et al., 2018; Himmelberg et al., 2021). This pattern suggests that retinotopic information does not end abruptly at the borders of early sensory areas, but rather fades progressively with cortical distance from V1. In other words, there may exist a macroscale gradient of functional specialization, consistent with classical models of hierarchical sensory processing.

We started from the idea that retinotopic selectivity might simply reflect distance from where visual thalamic input first arrives in the cortex, namely V1. To test this idea, we fit a threshold-linear model in which distance to V1 on the surface was the only predictor of selectivity (Figure 2.1). Even with this minimal approach, the geometric model explained 23.8% of the variance in the left hemisphere and 20.2% in the right. For comparison, the explainable variance based on the group-average selectivity map — an empirical upper bound — was 74.1% and 70.6%, respectively. This means that proximity to V1 alone accounts for a substantial share of the spatial structure in the data, without explicitly modeling connectivity or hierarchical relationships. The results reveal a clear spatial principle: retinotopic selectivity decreases gradually with distance from V1, and the decline extends well beyond early visual cortex. Areas such as V2 through V4 lie within 20 mm of V1 (Figure 1A,B), yet the gradient continues up to 60 mm along the cortical surface. This smooth decay shows that visual information is not restricted to isolated sensory modules but embedded within a large-scale, continuous gradient. Although the threshold-linear model based on geodesic distance captured the main organizational trend, it left structured residuals. Certain regions consistently showed higher or lower selectivity than predicted by distance alone (Figure 1D). These systematic deviations suggest that other spatial constraints or organizing forces are shaping the retinotopic landscape, pointing to the need for a more refined model.

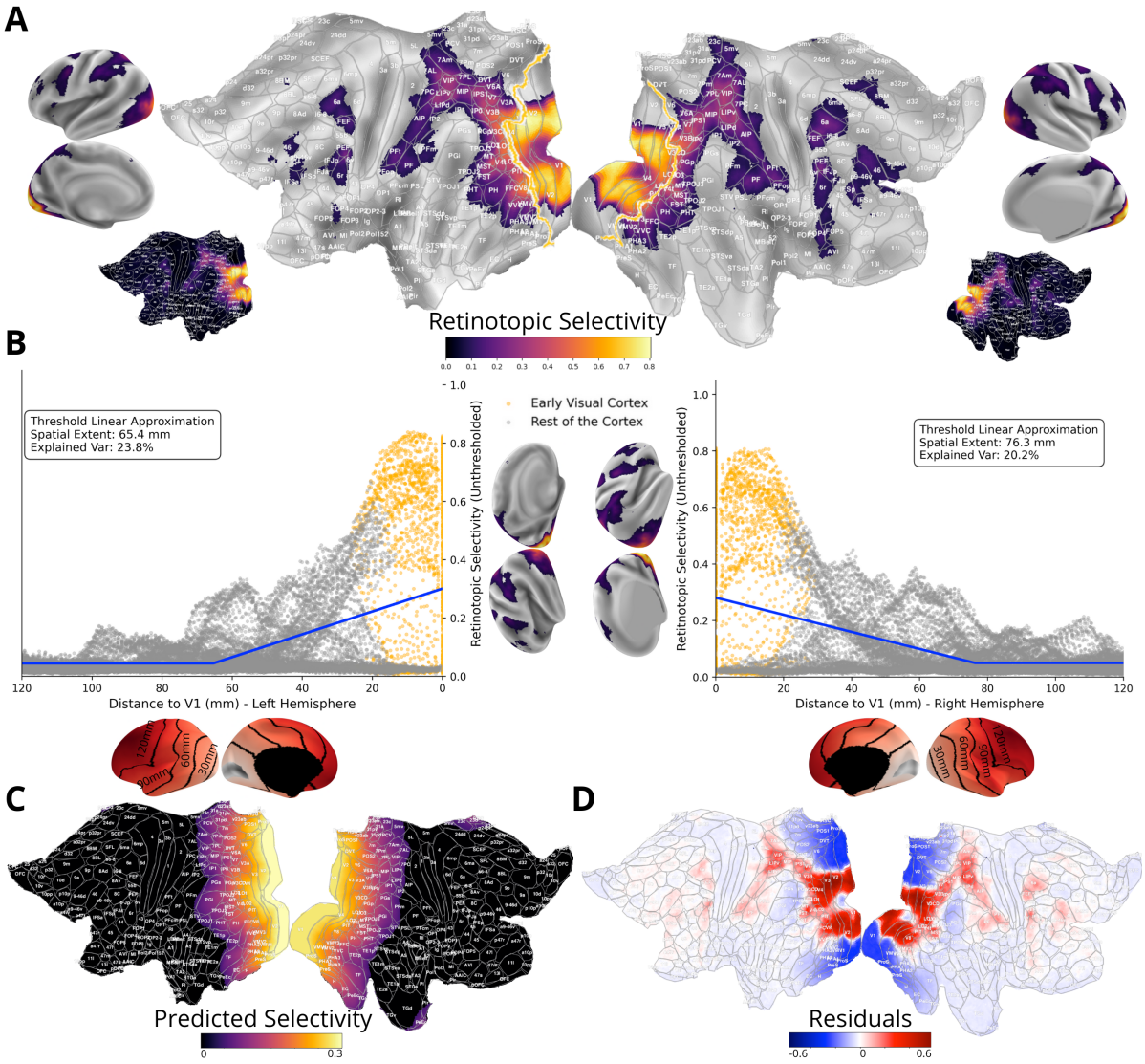

Figure 2.1: *Widespread retinotopic gradient beyond early visual areas.* ***A:*** *Retinotopic selectivity across the human cortex, shown on inflated and flattened surfaces for both hemispheres, with major anatomical landmarks overlaid. The yellow mask outlines early visual areas (V1–V4) based on the MMP1.0 atlas. A visualization ceiling was applied for contrast, with an unthresholded version shown in the bottom-left inset.* ***B:*** *Scatter plot of unthresholded retinotopic selectivity as a function of geodesic distance from V1. Yellow dots mark vertices within early visual areas (V1–V4); the blue line shows the threshold-linear model fit.* ***C:*** *Predicted selectivity from the threshold-linear model projected onto the flattened cortical surface with anatomical landmarks indicated.* ***D:*** *Model residuals on the cortical surface, highlighting regions that deviate from purely distance-based prediction.*

2.2 The Tethering Hypothesis

> *Everything is related to everything else, but near things are more related than distant things*
>
> Waldo Tobler

The macroscale gradient of retinotopic selectivity we just explored is not an isolated finding. It mirrors another key pattern from the previous chapter: a shift in NMDA-to-AMPA receptor density across the cortical surface. Together, these observations add to a growing body of evidence that the cortex is organized along large-scale axes, including variations in receptor density (Froudist-Walsh et al., 2023), myelination (Huntenburg et al., 2018; Paquola et al., 2019), or cell-type distribution (Kim et al., 2017). Such patterns don't need to be spatially smooth: they can emerge through coarse arealization, with regions adopting distinct profiles partly determined by their distance from key developmental anchors. During cortical morphogenesis, graded distributions of morphogens (Cadwell et al., 2019), or diffusible signaling molecules, guide the proliferation and differentiation of neural progenitors, thereby setting both the areal identity and laminar architecture of cortical regions. These biochemical gradients leave lasting spatial signatures in the distribution of neuron types, meaning that even anatomically discrete areas can participate in a broader, surface-based organizational framework. From this perspective, large-scale anatomical arrangements are not just empirical regularities but developmental fingerprints, lasting imprints of the processes that laid out the cortical sheet. Perhaps unsurprisingly, the identification of structural gradients was soon followed by the discovery of macroscale functional gradients—spanning dimensions such as intrinsic timescales (Murray et al., 2014) and functional connectivity profiles (Margulies et al., 2016). One interpretation is that these functional gradients emerge from underlying structural constraints. This is the central claim of the protomap hypothesis (Rakic et al., 2009), which posits that early molecular and cellular gradients preconfigure the functional roles of cortical regions. According to this view, the structural blueprint established during development constrains the eventual specialization of each area. An alternative perspective, captured by the protocortex hypothesis (Cadwell et al., 2019), suggests that the cortex begins as a largely undifferentiated sheet, with regional specialization emerging later through activity-dependent mechanisms. Given that cortical connectivity is predominantly local—with the probability of connection falling off exponentially with distance (Ercsey-Ravasz et al., 2013)—activity tends to propagate as neural waves across the surface (Muller et al., 2018; Pascucci et al., 2025). These waves may shape local circuits through Hebbian plasticity and even influence structural features over time, based on patterns of input. Today, the prevailing view integrates both perspectives. There is strong empirical support for early patterning mechanisms (e.g., morphogen gradients) as well as for activity-dependent refinement. Regardless of their relative weight, both frameworks

converge on a common outcome: cortical functional specialization emerges through macroscale gradient-like processes, whether driven by intrinsic structure, extrinsic input, or both.

One of the most elegant and provocative formulations of the protocortex idea comes from Buckner and Krienen (Buckner and Krienen, 2013). They argue that the spatial layout of the cortex is shaped by constraints imposed by thalamic sensory inputs. Specifically, they identify three cortical anchor points where focal thalamic projections deliver dense, structured input: V1 for vision, S1 for somatosensation, and A1 for audition. These primary sensory regions, by virtue of their direct coupling to the thalamus, are highly constrained in their architecture and function. But the influence of these anchors extends beyond the primary areas themselves. Surrounding regions—encompassing early and mid-level sensory cortices—are also shaped by their proximity to continuous streams of incoming sensory activity. Together, these zones are tethered: their local circuitry and functional specialization are deeply influenced by the sensory information they are exposed to. The real insight comes when this spatial logic is extended across the cortical sheet. The human cortex has expanded disproportionately in association areas—regions located far from these sensory anchor points. Buckner and Krienen suggest that this distance confers freedom: as cortical territories grow further from the gravitational pull of sensory input, they become progressively less constrained and more capable of supporting internally generated activity. These untethered regions are not defined by stimulus-driven processing; instead, they engage in abstraction, memory recall, and theory of mind—hallmarks of high-level cognition. This "tethering hypothesis" is one of my favorite theories of cortical organization. Not only does it capture the spatial asymmetries of cortical expansion, it also links development, anatomy, and cognitive function through a strikingly parsimonious principle: regions close to sensory input are shaped by it, while those farther away are free to wander the cognitive space.

This idea was extended by Margulies, Smallwood, and colleagues (Smallwood et al., 2021; Margulies et al., 2016), who proposed that the entire cortex may be organized, at least as a first-order approximation, by its distance from primary sensory anchors. Of course, cortical function is not determined by geometry alone—many other factors, from connectivity to gene expression, play critical roles. But as a starting point, cortical distance offers a remarkably powerful organizing axis. As the distance from primary sensory areas increases, neurons tend to respond less to raw sensory features and more to abstract, higher-order properties. This gradual shift in tuning resembles the hierarchical transformations seen in deep learning models, where successive layers move from encoding simple inputs to representing complex, integrated patterns. At the farthest extremes of this gradient lies the apex of the cortical hierarchy, where the default mode network acts as a multimodal integration hub, capable of manipulating highly abstract and internally generated representations. What makes this model so compelling is its simplicity. In principle, one could take a group of neurons on

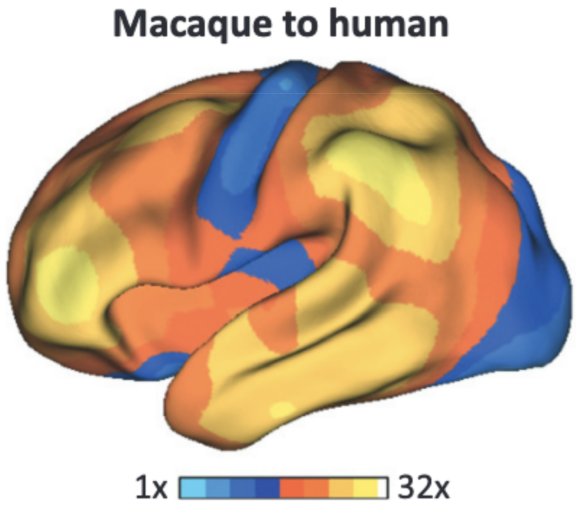

Figure 2.2: *Cortical expansion from macaque to human. Surface map showing the expansion factor of cortical areas when scaling from macaque to human brains. The color scale indicates relative expansion (blue = least expanded, yellow/white = most expanded), ranging from approximately 1× (blue) to 32× (yellow/white). Primary sensory areas show the lowest expansion, whereas association cortices display the highest relative expansion. Adapted from Buckner and Krienen 2013.*

the cortical surface, measure their distance to primary sensory regions, and use that distance to predict their functional specialization: what kind of computations they support, and what kinds of representations they are likely to encode. This offers a spatially grounded, biologically interpretable framework for understanding cortical organization in terms of both structure and function.

Describing the functional organization of the cortex in terms of spatial influences over a two-dimensional cortical sheet is, to me, a deeply appealing idea. It offers a unifying language to link structure, function, and development. But as we've already seen, capturing this organization is far from straightforward. In the relatively simple case of retinotopic selectivity, distance to V1 offers a reasonable first-order approximation of the map, yet leaves much of its structure unexplained. If a system as fundamental and straightforward as retinotopic selectivity resists simple spatial modeling, more complex functional architectures will require richer tools. This motivates the development of new methods to formally decompose and model spatial influences on the cortical surface.

2.3 Spatial Component Decomposition

> *We have the power to analyze and explore the world. But we panic when the data threaten not to validate our expectations*
>
> Franz de Waal

In recent years, several methods have successfully decomposed cortical maps into spatially structured components (Luppi et al., 2023; Pang et al., 2023; Vohryzek et al., 2025). One prominent example comes from Pang and colleagues (Pang et al., 2023), who demonstrated that task activation patterns could be reconstructed with high accuracy using eigenmode decompositions derived from the geometry of the cortical sheet. This approach was further extended by Vohryzek and colleagues (Vohryzek et al., 2025), who showed that adding a small set of long-range connections to the geometric eigenmodes significantly improved reconstruction performance. These methods compellingly demonstrate the role of cortical geometry as a scaffold for functional organization, and they highlight the spatial constraints shaping activity patterns across the brain. However, despite their success in reconstruction, these approaches rely on hundreds of eigencomponents, offering little anatomical specificity or interpretability. The resulting decompositions can reproduce the data well but do not directly yield insights into why the functional maps take the forms they do. Inferring underlying mechanisms—developmental, structural, or dynamic—from such decompositions remains an exciting challenge.

Building on these promising results, we developed a novel approach aimed at decomposing cortical maps into local spatial influences with explicit anatomical grounding(Equation 2, Figure 2.3). The core idea is simple: each influence originates from a single cortical location and decays linearly with distance across the cortical surface. These local fields can either enhance or suppress nearby map intensity, acting as positive or negative sources. Positive sources elevate functional values in their surrounding area, while negative sources reduce them. However, the total functional value at any cortical point is constrained to remain non-negative, meaning that negative sources can only diminish existing signals and have no effect in regions where no positive influence is present. This formulation extends the logic of the threshold-linear model introduced earlier, allowing for interactions between sources: a strong negative influence can locally cancel out the effect of nearby positive fields. By combining a small set of sources, the model seeks to reconstruct complex cortical maps in a way that is both compact and interpretable. The goal is not just accurate reconstruction, but explanatory clarity. Revealing where influential zones are located, how they interact, and eventually, what mechanisms might give rise to the observed functional organization.

$$Y(r) = \begin{cases} \sum_{j=1}^{P} a_j^p \, \phi_j^p(r) - \sum_{k=1}^{N} a_k^n \, \phi_k^n(r) & \text{if } Y(r) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

With Y denotes the target statistical map for all vertices r, ϕ^p and ϕ^n being respectively the facilitative and suppressive components, and a^p and a^n being the components loadings. Each component is parameterized as a threshold-linear function, as defined in Equations 2 and 3.

$$\phi_j^p(r) = \begin{cases} \beta_j \left(1 - \frac{d(r_j^p, r)}{\eta_j}\right) & \text{if } d(r_j^p, r) < \eta_j \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

$$\phi_k^n(r) = \begin{cases} -\beta_k \left(1 - \frac{d(r_k^n, r)}{\eta_k}\right) & \text{if } d(r_k^n, r) < \eta_k \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

With $d(r, r')$ denoting the geodesic distance between cortical vertices r and r'. Here, η defines the spatial extent of the component, β scales its maximum amplitude, and r^p and r^n specify the cortical locations of the positive and negative sources of influence, respectively.

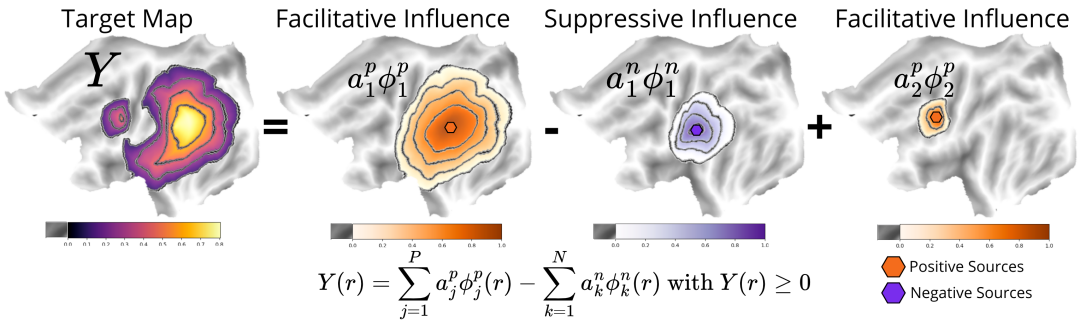

Figure 2.3: *Target map broken into spatially localized facilitatory and suppressive components.*

To formalize this decomposition, we implemented it within a sparse dictionary learning framework (Dohmatob et al., 2016; Varoquaux et al., 2011). This machine learning approach seeks to represent a high-dimensional signal as a weighted combination of a small number of elementary components or "atoms." The solution is constrained by an L0 norm penalty, which strongly discourages the inclusion of additional atoms and promotes compact, interpretable representations. Like LASSO regression, the

goal is to balance explanatory power with sparsity, but unlike convex regularization methods, sparse dictionary learning requires heuristic procedures to iteratively build the dictionary. In our case, the atoms are spatial influence fields—threshold-linear functions of geodesic distance from a candidate cortical vertex. Each atom has two free parameters: its peak magnitude at the source and its spatial extent. We added an additional biological constraint to the model: all predicted values must remain non-negative. This constraint reflects the empirical structure of the maps we aim to model (e.g., retinotopic selectivity), but it also complicates the optimization, making classic heuristics such as matching pursuit inapplicable. To solve this, we developed a custom greedy heuristic that takes advantage of the cortical surface's topological structure(Figure 2.4,algorithm 1). The algorithm begins by identifying the single best-fitting positive source—i.e., the vertex whose influence, with optimized parameters, explains the most variance on average across individuals. It then iteratively adds new sources: at each step, it selects the additional vertex that maximally increases explained variance, then allows all existing sources to be fine-tuned (in both location and parameters) to better account for shared variance. This cycle continues until no new source increases total variance explained by more than 0.1%. Because spatial influences are hemisphere-specific and do not cross the midline, the procedure is run independently for each hemisphere.

We applied this method to retinotopic selectivity maps from the Human Connectome Project (HCP) (Benson et al., 2018), which includes surface-based data from 181 participants. The dataset was split into training and test sets, and spatial influences were learned on the training data. To ensure generalizability and prevent overfitting to dataset-specific features, we cross-validated every spatial source using an independent dataset—the NYU replication dataset (Himmelberg et al., 2021), which includes 45 additional subjects. This procedure yielded a set of eight reproducible spatial influences per hemisphere, located at remarkably symmetric cortical positions(table 2.1,table 2.2).

Seed Vertex (fs_lr32)	Atlas (Yeo17 / Glasser)	Influence Type	Extent (mm)	Magnitude
23909	VisCent / V1	Facilitative	59.0	1.00
23864	VisCent / V1	Suppressive	19.3	0.79
25760	VisPeri / V1	Suppressive	30.1	1.00
14603	DorsAttnA / LIPv	Facilitative	45.4	0.44
16237	DefaultB / PGs	Suppressive	30.5	0.40
7436	SomMotA / Area 1	Suppressive	44.6	0.38
25149	VisPeri / V2	Suppressive	18.7	0.80
25272	VisPeri / V2	Suppressive	19.4	0.63

Table 2.1: *Spatially localized components identified in the left hemisphere using the HCP dataset.*

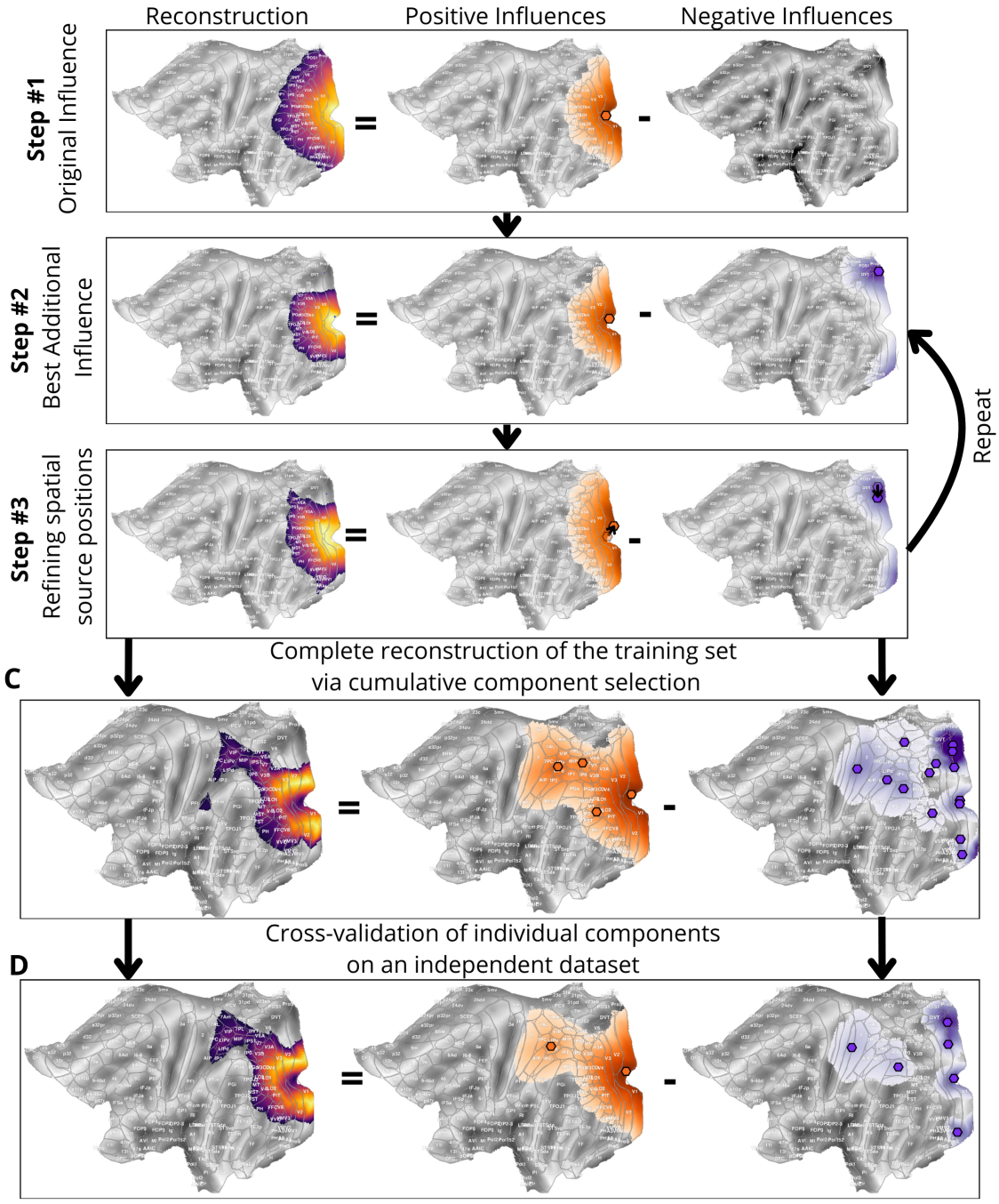

Figure 2.4: *Heuristic of the Spatial Component Decomposition*

Algorithm 1 Greedy Heuristic for Sparse Spatial Decomposition with Validation

Require: Training data Y_{train}, Test data Y_{test}, Validation data Y_{val}, threshold λ, significance level α
Ensure: Final set of validated atoms $\mathcal{A}^*$
Initialize empty atom set $\mathcal{A} \leftarrow \emptyset$
while True **do**
    for all candidate vertices v **do**
        Form tentative atom A_v centered at v
        Jointly fit amplitudes for $\mathcal{A} \cup \{A_v\}$ on Y_{train}
        Evaluate total variance explained
    end for
    Select $A^* = \arg\max_{A_v}$ (variance explained on Y_{train})
    Jointly refit amplitudes for $\mathcal{A} \cup \{A^*\}$ on Y_{train}
    Evaluate improvement ΔV on X_{test}
    if $\Delta V < \lambda$ **then**
        break
    end if
    Add A^* to $\mathcal{A}$
    Local refinement:
    for all $A_i \in \mathcal{A}$ **do**
        while neighboring move improves fit on Y_{train} **do**
            Move center of A_i to neighboring vertex
            Refit amplitudes for all atoms in $\mathcal{A}$
        end while
    end for
end while
Cross-dataset validation:
Fix atom locations in $\mathcal{A}$
Refit amplitudes and spatial extents on Y_{val}
for all $A_i \in \mathcal{A}$ **do**
    Compute full model performance V_{full}
    Compute reduced model performance V_{-i} without A_i
    Generate null distribution by shuffling model labels assigned to subject-wise losses in Y_{val}
    Compute p-value as the proportion of permutations where the mean loss difference exceeds the observed difference
    if $p < \alpha$ **then**
        Retain A_i
    else
        Discard A_i
    end if
end for
Let $\mathcal{A}^* \subseteq \mathcal{A}$ be the retained atoms
return $\mathcal{A}^*$

Seed Vertex (fs_lr32)	Atlas (Yeo17 / Glasser)	Influence Type	Extent (mm)	Magnitude
23949	VisCent / V1	Facilitative	60.4	1.00
23907	VisCent / V1	Suppressive	20.1	0.83
26109	DefaultC / POS1	Suppressive	50.0	1.00
14572	DorsAttnA / LIPv	Facilitative	56.9	0.62
6804	SomMotA / Area 2	Suppressive	50.0	0.49
16123	ContB / PGs	Suppressive	35.3	0.40
25595	VisPeri / V1	Suppressive	30.9	1.00
25005	VisPeri / V2	Suppressive	14.8	1.00

Table 2.2: *Spatially localized components identified in the right hemisphere using the HCP dataset.*

Our Spatial Component Decomposition explained 68.4% of the variance in the left hemisphere and 65.2% in the right when reconstructing retinotopic selectivity. Relative to the explainable variance—defined by the group-average map in the training set (74.1% and 70.6%, respectively)—this corresponds to 92.3% of the total predictable signal. Notably, this performance was achieved with only eight spatial components per hemisphere, amounting to just 24 free parameters. This compact representation suggests that much of the cortex's retinotopic topography can be reconstructed from a small number of anatomically grounded spatial influences. For comparison, we implemented a connectivity-based model using structural tractography data from the HCP, which provides vertex-level estimates of structural connectivity across the cortical surface. Using the same sparse dictionary learning framework and matching the number of parameters, the connectivity-based model explained only 63.9% of the variance in the left hemisphere. This highlights the superior predictive power of local cortical geometry over long-range structural connections for this particular functional property.

Just as Vohryzek and colleagues enhanced the performance of geometric eigenmode reconstructions by incorporating a sparse set of long-range connections (Vohryzek et al., 2025), future extensions of our model could similarly benefit from allowing sparse non-local interactions—potentially improving reconstruction accuracy while retaining interpretability.

In principle, Spatial Component Decomposition can be applied to any cortical map to reveal the spatial influences that shape it. It remains, of course, a modeling tool, and the influences it highlights will ultimately require confirmation through mechanistic investigation. We plan to release a full implementation of the method to enable its use on other datasets. One important caveat is that the decomposition can be unstable, particularly for smaller components: slight changes in the selection procedure may produce markedly different solutions and introduce non-generalizable

spatial influences. In our own analysis, the initial set of 18 components was reduced to 8 after cross-validation, underscoring the need for stringent validation. In the present case, the availability of an external dataset allowed us to mitigate this limitation, but the issue remains a weakness of the approach. Preliminary experiments suggest a possible path toward greater robustness. Using a Monte Carlo selection procedure, in which additional sources are chosen probabilistically via a softmax function of the variance they explain, and the process is repeated many times, yields more stable decompositions. Although still at an early stage, this approach offers a promising direction for improving the generalizability of Spatial Component Decomposition.

2.4 Revealing Structured Spatial Influences on Retinotopic Maps

Applying Spatial Component Decomposition to the Human Connectome Project retinotopic selectivity maps revealed a small set of highly reproducible spatial influences. Eight components per hemisphere were identified, each localized to a distinct cortical region and exhibiting a consistent anatomical pattern across individuals. Strikingly, the arrangement was bilaterally symmetric, with corresponding sources appearing in homologous locations in left and right hemispheres. In both hemispheres, the most prominent component was centered in V1, in line with the idea that thalamic input (Callaway, 2005) is the primary driver of retinotopy on the cortical surface. Remarkably, the spatial reach of this single influence extended over 60 mm, well beyond the boundaries of primary visual cortex. This alone shows that much of the visual cortex's retinotopic selectivity can be accounted for by a single, dominant source. It could have been otherwise. Each visual area might have carried its own local source of retinotopic selectivity, perhaps decreasing gradually in strength along the visual hierarchy. Instead, the model favored a simpler explanation: one large, positive influence emanating from V1, radiating across multiple visual areas. Complementing this dominant positive influence, the decomposition also revealed two smaller, negative sources within V1 in each hemisphere—one located on the occipital pole and the other at the most anterior part of the calcarine sulcus. To better understand the origin and organization of these spatial influences, we begin by examining those arising from V1. Interestingly, this area contributes not only a dominant facilitatory gradient but also localized suppressive effects(Figure 2.5).

V1 is arranged topographically along an antero–posterior axis: the central visual field is represented in the occipital portion of the calcarine sulcus, while the far periphery is mapped to its most anterior part (Figure 2.5) (Wu et al., 2012; Haenelt et al., 2023). One of the two negative sources identified by our decomposition falls within this eccentric region of the visual field. This finding can be readily explained by the experimental design of the Human Connectome Project data. Because of stimulus

constraints in the fMRI protocol, the visual stimuli never extended beyond 16 degrees of eccentricity, whereas the human visual field reaches about 60 degrees. Since our measure reflects the fraction of neural activity explained by the position of the probe within retinotopic space, it is expected that these far-peripheral zones would show little or no selectivity—they simply lie outside the stimulated range. The more surprising observation is the size of this negative influence. Its spatial extent reaches well beyond the borders of V1, implying that these regions are not only shielded from the main gradient originating in V1 but may also exert a suppressive effect on the surrounding cortex.

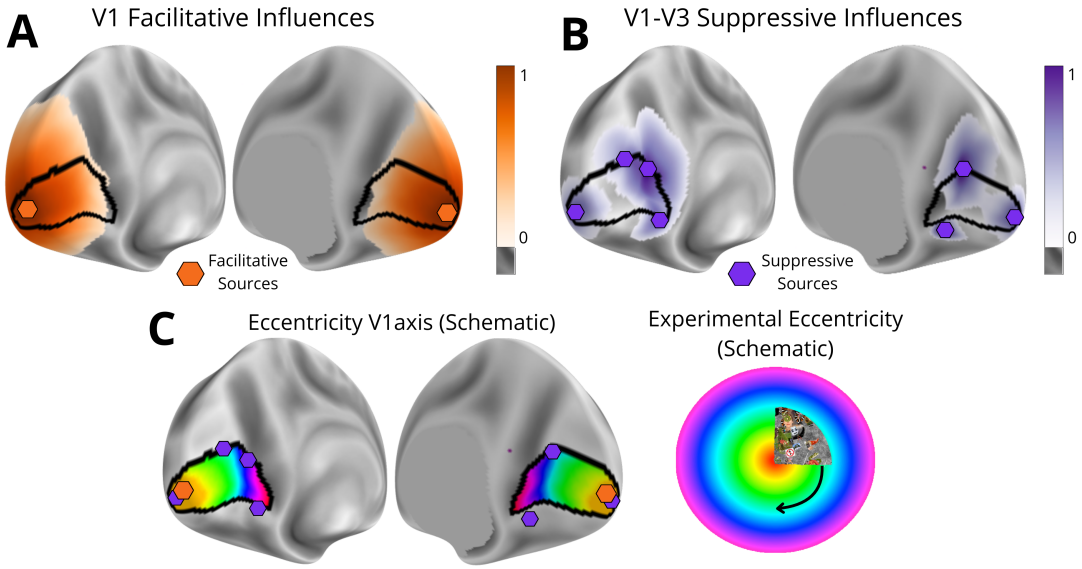

Figure 2.5: *Sources localized in early visual cortex (V1–V3).*

The second negative source was more unexpected. It lies at the very center of the retinotopic field, a region that was fully stimulated during the experimental protocol. When we examined the functional connectivity of these sources using the HCP resting-state dataset (Van Essen et al., 2013), we found that although their functional connectivity was lower than the cortical average, their activity correlated strongly with transmodal regions such as the default mode network and the control network, as well as with primary visual areas(Figure 2.6A). This raises the intriguing possibility that these regions serve as a bridge between higher cognition and sensory processing, particularly for information at the center of our gaze—content that may receive enhanced cognitive resources. Interestingly, this finding also resonates with recent reports of transmodal focal regions embedded within primary somatosensory cortex. To explore this effect further, we applied k-means clustering to the vertices of V1 in each hemisphere, using their functional connectivity to the rest of the cortex as features(Figure 2.6B,C). The results revealed a striking bimodal pattern: a small polar

cluster, corresponding closely to our negative sources and showing strong connectivity with transmodal regions, and a much larger cluster covering the remainder of V1, which was primarily coupled to other primary sensory cortices and largely decoupled from transmodal areas (Figure). When we projected the V1 vertices into the connectivity space described by Margulies et al. (2016) (Margulies et al., 2016), which captures the two principal axes of cortical functional organization, the polar cluster occupied a position near the transmodal regions(Figure 2.6D). This reinforces the view that this portion of V1 has a distinctive connectivity profile that sets it apart from the rest of the area.

A systematic analysis of the functional connectivity of all negative sources in each hemisphere confirmed the pattern found within V1(Figure 2.7). These sources fell into two main categories. The first included other primary sensory regions, such as the eccentric portion of V1, as well as S1, which appeared in both hemispheres. The presence of S1 is notable, given that no strong somatosensory stimulation occurred during the retinotopy task. This reinforces the idea of a competition for cortical territory that operates independently of the immediate sensory stream. The second category consisted of transmodal hubs from the Yeo parcellation, in particular the default mode network. The DMN is often associated with internally directed cognition, including memory, theory of mind, and imagination (Buckner and DiNicola, 2019). This raises the intriguing possibility that internally generated representations also compete for space on the cortical surface, influencing sensory maps well beyond their traditional boundaries.

As with the V1 sources, we extended our analysis to the functional space described by Margulies and colleagues in 2016 (Margulies et al., 2016), systematically projecting each source onto the two principal connectivity gradients(Figure 2.8). In this representation, the cortex takes on a roughly triangular shape, with each vertex corresponding to a primary sensory system: visual, somatosensory, and the transmodal default mode network. Strikingly, every single source—positive or negative—fell along the apex of this functional space. This supports the idea that these apex represent strong poles of functional specialization, exerting spatial influence over their immediate surroundings. In contrast, the center of the space appears to be occupied by more flexible, integrative regions that combine information from these specialized systems. This spatial antagonism between primary sensory regions and the default mode network may be another facet of the well-established temporal competition between them.

Finally, the only relay source that survived cross-validation was located in the lateral intraparietal area (LIP), within the dorsal attention network. Interestingly, additional positive sources were also identified in frontal regions belonging to the same network. Although these did not survive cross-validation—likely due to lower anatomical alignment in frontal cortex—they hint at a broader pattern: that the secondary

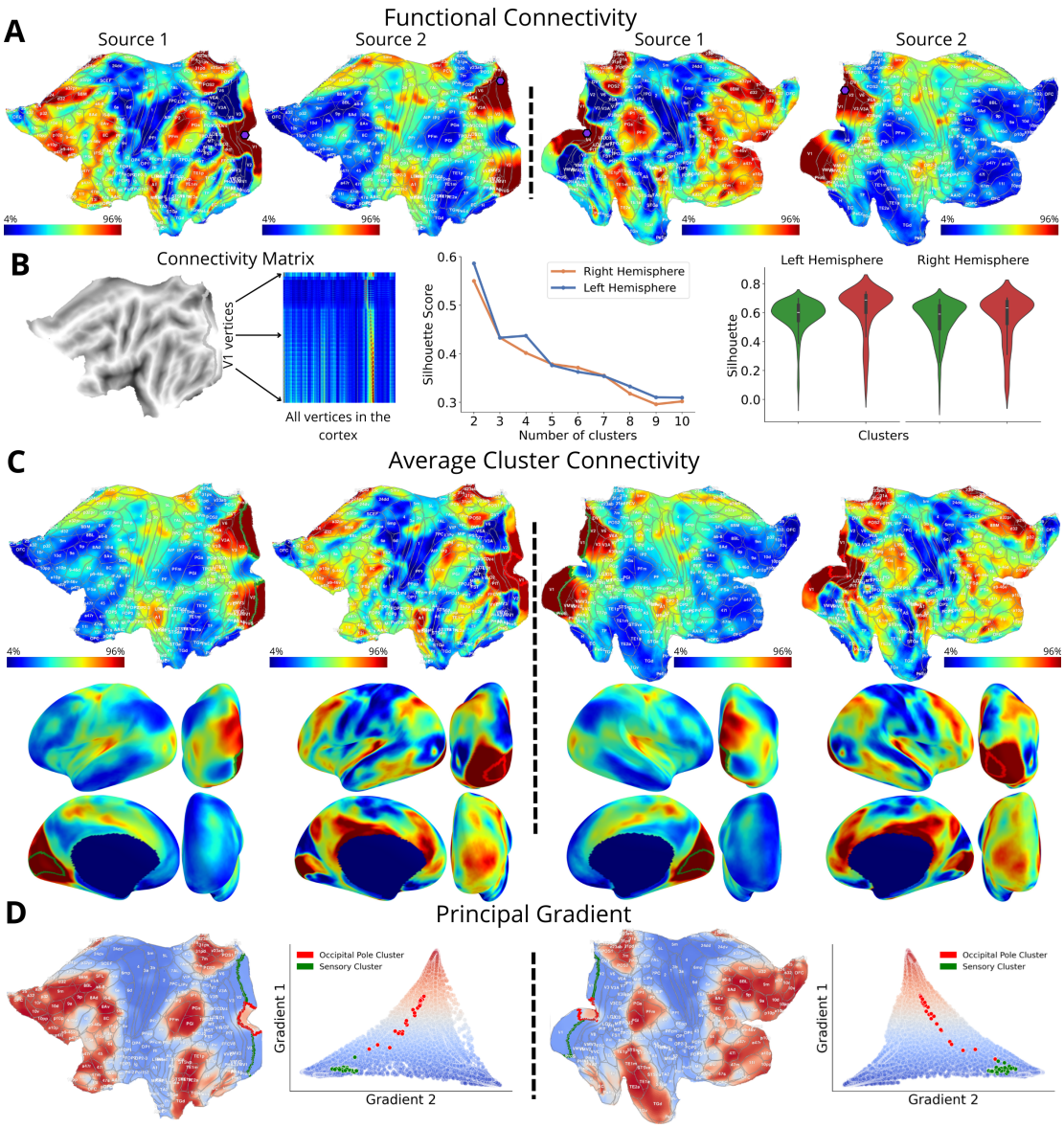

Figure 2.6: *Two distinct functional clusters within V1.* ***A*** *Functional connectivity maps for each of the two V1-based negative sources, showing distinct patterns of functional connectivity.* ***B*** *(left panel) Functional connectivity matrix between all V1 vertices (rows) and the rest of the cortex (columns), sorted by cluster assignment. (middle panel) Silhouette score as a function of the number of clusters used in k-means clustering. (right panel) Distribution of silhouette scores within the two selected clusters (bottom).* ***C*** *Average whole-brain functional connectivity profiles for each V1 cluster.* ***D*** *Principal functional connectivity gradient (G1) from Margulies et al. (2016), with the two V1 clusters overlaid in red and green.*

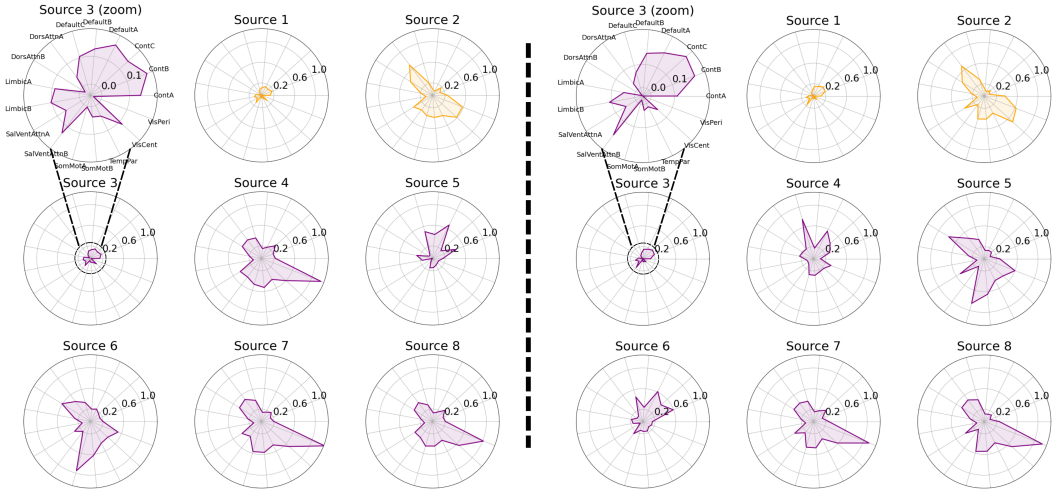

Figure 2.7: *Connectivity profiles to Yeo networks for all identified sources. Each spider plot represents the average functional connectivity from a given source to each of the seven Yeo networks.*

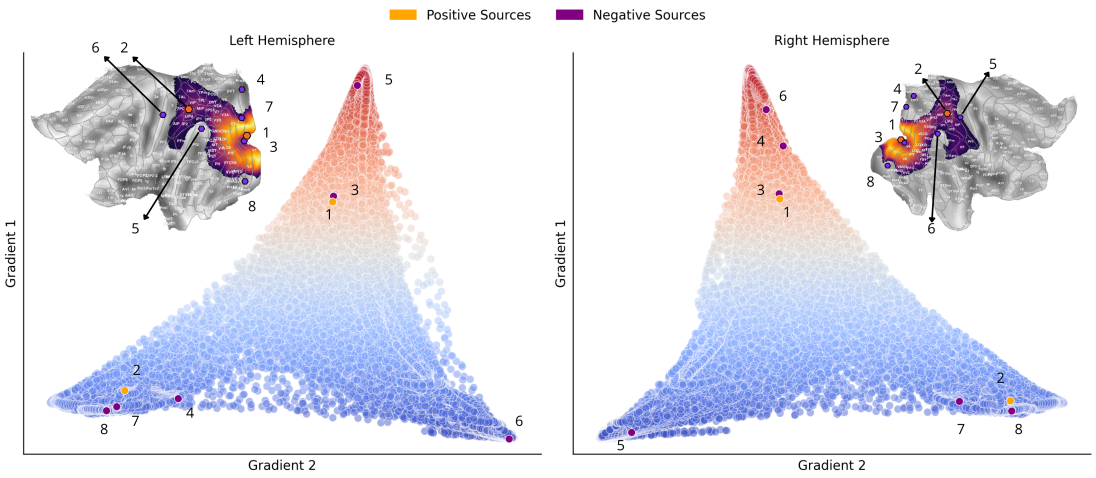

Figure 2.8: *Projection of all sources in functional connectivity space.*

facilitatory influences (excluding the dominant V1 source) are concentrated within the dorsal attention system. This observation is particularly intriguing given that LIP was also one of the key nodes in the bistable associative network identified in Chapter 1. It suggests that the dorsal attention system may serve not only as a relay but as an amplification hub, selectively enhancing and broadcasting relevant sensory signals across the cortex.

2.5 Open Questions and Future Work

These findings point to a broader organizing principle: mental representations appear to compete for space on the cortical surface, with strong sources shaping activity both locally and in surrounding regions. If this principle holds, it should apply not only to vision but also to other primary sensory modalities, as well as to higher-order, non-sensory representations.

We extended the spatial component decomposition to functional connectivity patterns seeded from the three primary sensory regions, following recent cortical models of multimodal sensory processing (Wei et al., 2024). While these decompositions have not yet been stabilized through external cross-validation, they consistently reproduced the key features observed in the retinotopy analysis: in each case, a strong positive source emerged within the corresponding primary sensory cortex. And in each case, the negative influences appeared in nearby default mode regions and in other primary sensory systems competing for cortical territory. We next turned to cortical maps from Doerig et al. 2025 (Doerig et al., 2025), which estimate for each vertex the similarity between its activity and the highest layers of large language models describing the viewed images. This metric served as our proxy for the functional specialization of higher sensory processing. Here, the dominant positive influence was centered on area MT, while systematic negative influences arose from adjacent default mode regions as well as from V1. Although these results are preliminary and not yet submitted for publication, they align with the idea of a general organizing principle in which strong representational sources—whether sensory or higher-order—engage in spatial competition.

It is important to recognize that the brain's encoding of information is not static—it likely shifts depending on cognitive state and the current focus of conscious awareness. In the task used to estimate retinotopic selectivity, participants fixated on a rotating wheel filled with salient and engaging images—objects, animals, and people—and their attention was fully directed toward these visual stimuli. This setup captures a state of active visual engagement. However, the resulting cortical representation would likely differ if participants were engaged in a different task—such as attending to sounds or performing a working memory challenge—while the same visual stimuli remained present but unattended. Indeed, studies show that retinotopic selectivity drops in

associative areas like LIP and prefrontal cortex when the stimulus is not the focus of attention, whereas early visual areas such as V1 are far less affected (Bressler and Silver, 2010). This suggests a hierarchical modulation of representation by conscious focus: primary sensory areas encode stimuli regardless of awareness, while higher-order relay regions selectively amplify what enters the stream of consciousness. A parsimonious explanation is that primary sensory cortices encode stimuli regardless of conscious access, largely because they receive direct influence from V1, which processes visual input whether or not it is consciously perceived. In contrast, higher-order relay centers like LIP become active primarily when the stimulus reaches conscious awareness, and they in turn propagate information to more distal associative areas. In this framework, cortical regions closer to V1 can be activated through bottom-up sensory input alone, while those closer to LIP rely more heavily on top-down relay—becoming engaged only when the stimulus is consciously accessed. This gradient of dependence naturally accounts for the gradual loss of selectivity observed across the cortical sheet, and offers a coherent bridge between the spatial geometry of representation and the dynamic gating mechanisms that underlie conscious perception.

3

Self-Organizing Cortex: Mechanisms and Proposed Architecture

*The work described in this chapter led to a poster presentation at the 24th Annual Computational Neuroscience Meeting (CNS*2015) on Hebbian receptive field formation, a successful grant proposal on reward-based self-organization (currently being developed), and a review manuscript on large-scale cortical organization, currently in preparation.*

In the previous chapter, we uncovered structured spatial influences that shape how mental representations are distributed across the cortical sheet. For retinotopic selectivity, the dominant positive influence aligns closely with the thalamic projections to visual cortex, accounting for much of the large-scale map. This influence is relayed and extended by dorsal attention regions, while being systematically counterbalanced by suppressive influences from other primary sensory cortices and transmodal hubs of the default mode network. One striking result is that other sensory regions exert a measurable influence despite being only weakly stimulated—or not stimulated at all—during the task. This suggests that their impact is not purely a consequence of moment-to-moment information flow, but may instead reflect a persistent trace embedded in the cortical architecture. Such influences could arise from long-term connectivity patterns, residual activity states, or stable representational structures that remain even in the absence of direct input. The logic of this arrangement—facilitatory streams radiating from primary entry points, selectively amplified by attentional systems, and suppressed by competing sensory domains and transmodal nodes—appears to generalize beyond vision. Preliminary observations suggest similar spatial patterns in other modalities and in higher-order representations, though these remain to be confirmed. This opens an important opportunity: to move from descriptive mapping toward a mechanistic account. If such spatial influences recur across modalities, what processes give rise to them? What anatomical and physiological constraints could generate both the long-range facilitatory gradients and the localized suppressive zones revealed by our analyses?

In this chapter, we build on these preliminary results to propose a general mechanism based on Hebbian plasticity, capable of generating both the facilitatory gradients and the suppressive zones revealed in our analyses. Building on this foundation, we also outline a tentative global architecture of cortical organization, aiming to explain how these spatial influences emerge and interact across the cortical sheet.

3.1 The Formation of Receptive Fields

One of the most famous statements in neuroscience came from Donald Hebb: "Neurons that fire together wire together." This deceptively simple principle captures a fundamental property of the brain—its connections are not fixed, but highly plastic, adapting to patterns of neural activity. In essence, the stronger the correlation between the spiking patterns of two neurons, the stronger the connection between them will tend to become. This core idea has been refined over decades since Hebb's original proposal, most notably through Spike-Timing Dependent Plasticity (STDP) (Caporale and Dan, 2008; Markram et al., 1997). STDP formalizes the relationship between the precise timing of neuronal firing and the resulting changes in synaptic strength. The logic is straightforward: if neuron A fires just before neuron B, the synapse from A to B is strengthened; if the order is reversed, the synapse is weakened—though typically to a lesser degree. Over time, this mechanism can give rise to distinct organizational motifs. In some cases, it creates directed chains of neurons, where synaptic strengths are aligned so that activity reliably flows from one cell to the next. In others, it produces tightly interconnected ensembles, in which every neuron shares strong bidirectional connections with the rest, enabling stable and self-sustaining activity. Conversely, pairs of neurons whose firing patterns remain uncorrelated will see their connections gradually weaken, eventually becoming functionally absent. In this way, Hebbian plasticity sculpts the network by reinforcing pathways of coordinated activity while pruning those that carry no consistent signal.

Together with Claudia Clopath and Jacopo Bono at Larry Abbott's lab, we explored the formation of receptive fields through Hebbian learning by simulating a spiking neural network (Clopath et al., 2015). This work extends earlier studies by Miller and colleagues (Miller et al., 1989), who demonstrated how Hebbian plasticity can give rise to ocular dominance columns in the primary visual cortex, by applying similar principles to a broader, spatially structured input regime(Figure 3.1). The network we built begins with random connectivity and received spatially structured input designed to mimic visual signals. Hebbian plasticity was paired with a homeostatic constraint (Clopath et al., 2010), ensuring that each neuron maintained roughly the same total amount of incoming connections over time, even as individual synapses were strengthened or weakened. The network also included inhibitory populations, which introduced competition between excitatory neurons and led to a sharpening of stimulus

representations. Although this work has not yet been peer-reviewed and has so far only been presented as preliminary results at a conference, it revealed dynamics highly relevant to our present argument. Over time, each neuron in the population developed a receptive field tuned to specific retinotopic locations and formed stronger connections with other neurons sharing similar preferences. This process produced a distinctive connectivity architecture in which, even in the absence of visual input, the intrinsic local connections preserved the topology of the previously experienced stimulus patterns (Clopath et al., 2015). Remarkably, this structural organization was mirrored in the network's spontaneous activity, yielding a functional connectivity map that aligned with the neurons' receptive field preferences. This pattern closely resembles what has been observed experimentally in the visual cortex. In other words, sensory activity left a lasting imprint on the local connectivity, persisting well beyond stimulus removal. At the local scale, such persistent structure could provide a mechanistic origin for the negative influences uncovered by the Spatial Component Decomposition in the previous chapter.

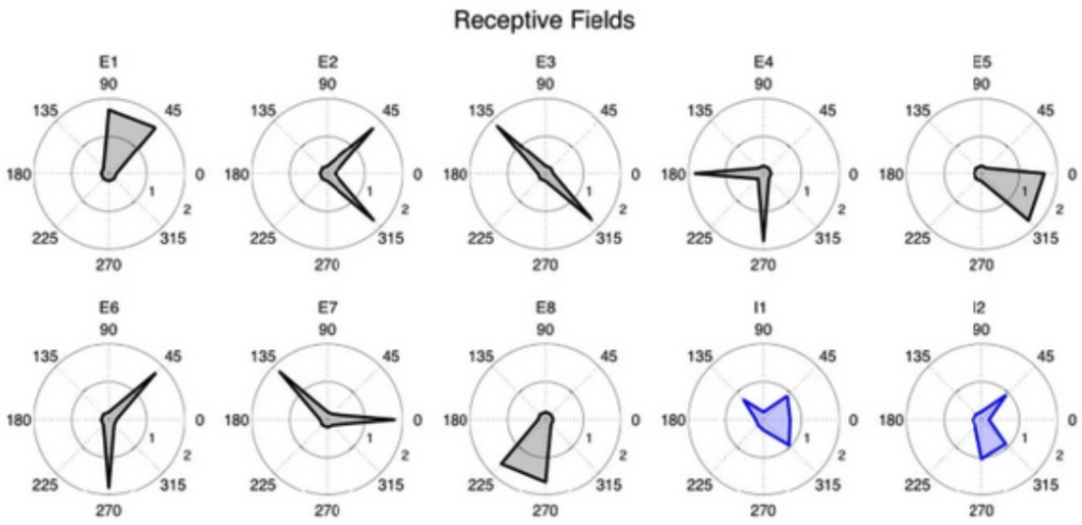

Figure 3.1: *Emergence of receptive fields through Hebbian plasticity during passive visual exposure.*

These results demonstrate that a population receiving spatially structured input can, through local plasticity, form connectivity patterns that define receptive fields. This mechanism could underlie the emergence of receptive fields in primary visual cortex (V1), which receives strong, topographically organized visual input from the thalamus. Notably, a substantial portion of receptive field selectivity in V1 develops even before eye opening, potentially driven by spontaneous waves of neural activity in the retina that are relayed to V1 (Martini et al., 2021). Together, these findings lend support to the protocortex view of functional specialization (Cadwell et al., 2019), at

least in the case of V1, where structured input and intrinsic plasticity appear sufficient to establish a stable representational architecture.

Once V1 acquires strong receptive fields dominated by retinotopic information, it acts as a structured input source for V2, transmitting an already organised map. This input can in turn drive the emergence of receptive fields within V2. The same local architecture, capturing retinotopic information and shaping preferred receptive field specializations, can continue to spread to any region receiving strong projections from V1 or V2. Because most cortical connections are local, and the probability of a connection between two neurons decreases exponentially with distance, this process could give rise to a large-scale gradient of connectivity. In such a gradient, retinotopic information would gradually diminish in strength, yet remain detectable far beyond the boundaries of the early visual cortex. It is important to acknowledge that these ideas remain speculative and will require formal investigation in future work. For this proposed mechanism to account for the observed gradient of retinotopic information, the local connectivity pattern would need to weaken at each step, leading to a progressive decrease in retinotopic selectivity as one moves away from V1.

An appealing aspect of this proposed mechanism is its potential generality. It could, in principle, apply to any form of structured input, whether somatosensory, auditory, or even internally generated mental representations, provided there is a consistent correlation structure among different percepts. The same process could therefore originate from S1 (Kaas, 1993), A1 (Hackett, 2011), or any other cortical region receiving strong, topographically organised input, such as the entorhinal cortex during hippocampal replay (Witter and Amaral, 1991). If the mechanism holds, each of these representations would be expected to propagate across the cortical surface, gradually extending into neighbouring regions and reconfiguring local connections to capture the topology of their respective representational domains. We can push the speculation further. As different representations spread across the cortical surface, they will eventually collide. In the regions where they converge, neurons will receive inputs from two largely uncorrelated input streams, creating a form of representational conflict. Each neuron could then develop two distinct preferred receptive fields, one for each modality. For example, a region at the intersection of visual and somatosensory streams might contain neurons tuned both to a specific foveal location in retinotopic space and to a particular location on the hand (Figure). Such neurons would still tend to connect locally with others sharing either receptive field, but the strength of these connections would be reduced, since the two modalities are only weakly correlated(Figure 3.2). There are many situations in which visual input occurs in the foveal field without any concurrent stimulation of the hand, and vice versa. Pushing this idea further, because of Hebbian plasticity, the connectivity of regions at the crossroads of two modalities will be shaped by the statistical regularities in their combined inputs, even when those inputs come from different sensory channels. In such cases, the tuning of neurons

could be driven by the covariance between the two signals, effectively implementing multimodal integration. This idea is supported by the lateral intraparietal area (LIP), which sits at the intersection of the visual and somatosensory streams and contains neurons with receptive fields in both modalities (Sereno and Huang, 2006; Sood and Sereno, 2016). Remarkably, these receptive fields are not independent but aligned: when the face is touched, the perceived direction of that tactile stimulus is represented in a way that corresponds spatially to the neuron's visual receptive field, as revealed by population receptive field mapping (Sereno and Huang, 2006).

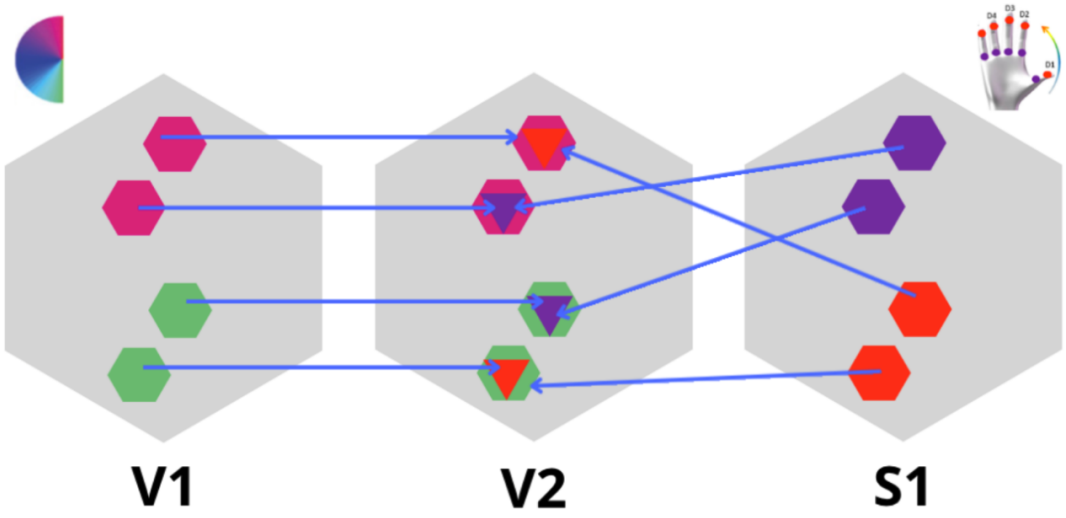

Figure 3.2: *Schematic of receptive field formation from multimodal input. Illustration of how a neuron or population can develop distinct receptive fields in response to two spatially structured inputs from different sensory modalities, via competitive Hebbian learning.*

With these speculative scenarios in mind, we can now step back and consider their broader implications. Based on the discussions above, the spatial influences identified in the previous chapter could reflect the local reconfiguration of connectivity, encoding different forms of mental representation. In this view, which aligns with the protocortex perspective on functional specialization, the cortical sheet becomes a competitive landscape where different information streams expand their territory until they meet, at which point their interaction can give rise to new representational content. This framework could also explain why the transmodal regions of the default mode network appeared as negative sources in our decomposition. These regions may serve as origins of internally generated representations, such as theory of mind or memory replays, which compete for cortical territory alongside sensory streams. This interpretation is compatible with the tethering hypothesis, but with a nuanced extension: transmodal hubs would also act as spatial anchors on the cortical surface,

their separation from primary sensory regions facilitated by cortical expansion. This separation could both free these hubs from sensory constraints and shield primary sensory areas from internally oriented activity, ensuring that higher-order cognition does not come at the expense of sensory processing.

The combination of two simple rules, specifying entry points for information on the cortical surface followed by the propagation and refinement of local connectivity, could provide an elegant account of how the cortex self-organizes. In this view, the dorsal visual stream, extending from V1 through the intraparietal sulcus, would occupy its characteristic location partly to avoid the neighbouring default mode network nodes of the angular gyrus and the posterior medial cortex. It would eventually converge with the somatosensory stream in the LIP region, where multimodal integration could occur. Such an arrangement could arise purely from the self-organization of functional architecture driven by the structure of incoming signals.

Yet this picture leaves out an important dimension of brain plasticity: learning shaped by reward. Dopaminergic signals can trigger powerful changes in connectivity (Jay, 2003; Reynolds and Wickens, 2002), suggesting that elements of the brain's cognitive architecture may arise through a different mechanism. When a sequence of neural activity culminates in a rewarding event, such as the successful identification of food, the synaptic connections that contributed to that sequence are selectively reinforced. This principle lies at the heart of the backpropagation framework widely used in machine learning (Rumelhart et al., 1986), which has produced remarkable advances in the cognitive abilities of artificial neural networks over the past decade. Although the brain is unlikely to implement backpropagation in its exact mathematical form, it may rely on alternative algorithms that converge toward similar connectivity patterns (Whittington and Bogacz, 2019; Song et al., 2021).

We explored this idea in collaboration with Adrien Doerig by embedding artificial neural networks onto the cortical surface, placing a unit at every location of the fs_lr32 mesh. The probability of a connection between any two artificial neurons depended on their physical distance, following an exponential distance rule. Inputs were injected into the V1 region, and outputs were read from the LIP region. The network was trained using backpropagation, and after learning, we observed the emergence of neural activity propagating along the intraparietal sulcus, precisely in the location corresponding to the biological dorsal visual stream. These preliminary results suggest that the backpropagation rule is also a strong candidate for driving the self-organization of cortical functional architecture. Future work will examine these ideas in greater depth, particularly by incorporating multimodal stimuli and more complex behavioural tasks. The resulting picture is that the functional architecture of the brain may arise from a combination of two self-organization principles: the competition for cortical territory between different forms of information through Hebbian learning, and the reinforcement of the most relevant information through reward-based plasticity.

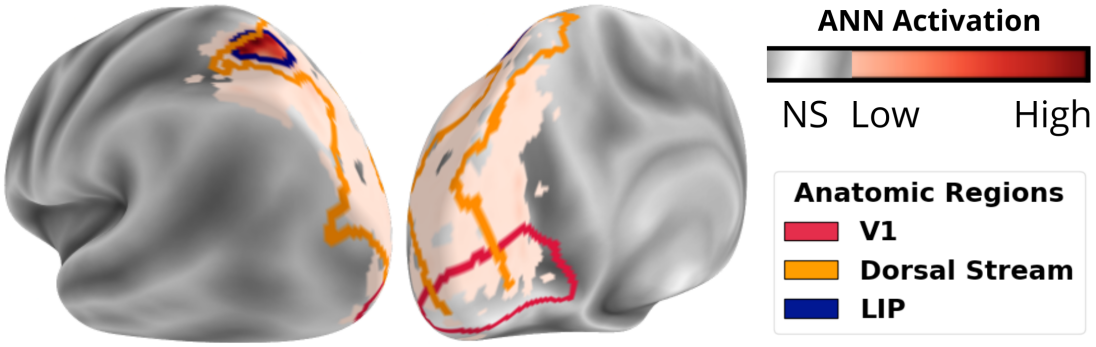

Figure 3.3: *Preliminary results – Emergence of the dorsal visual pathway in our model. Model activity plotted on the cortical surface shows alignment with known dorsal stream regions, supporting the emergence of functional pathways.*

3.2 Global Architecture Emerging from Spatial Competition

The retinotopic selectivity map traces a clear and familiar path. It follows the dorsal visual stream (Scolari et al., 2015; Szczepanski et al., 2013). High selectivity values appear first in early visual areas, including V1, V2 and V3. These regions are organized in concentric retinotopic maps, with each area encircling the previous one, reflecting a systematic and hierarchical encoding of visual space. From there, they continue along the intraparietal sulcus and eventually reach the anterior parts of the parietal cortex, notably LIP and VIP. This pattern closely matches the visual hierarchy architecture first described by Van Essen and Felleman (Felleman and Van Essen, 1991). In their framework, primary regions handle basic visual features, while the intraparietal sulcus helps extract spatial information progressively. LIP then transmits this processed signal toward frontal regions through strong white matter connections. But when we overlay this classical view with the spatial influence patterns from the previous chapter, a new story begins to take shape. The signal appears to originate in V1, where thalamic input first reaches the cortex, and then moves across the cortical surface. As it progresses, it is shaped by suppressive forces from two transmodal hubs, the angular gyrus and the postero-medial cortex, both part of the default mode network. These regions flank the intraparietal sulcus and seem to constrain the spread of retinotopic information, forcing it through a narrow functional corridor. At the end of this path lies LIP, which acts as a relay, reamplifying the signal and sending it toward the frontal cortex. Remarkably, V1 and LIP are also the only two significant positive sources identified in the Spatial Component Decomposition which survived the cross validation. Additional sources were observed within the attention network—particularly in frontal regions—but did

not consistently survive validation, likely due to greater inter-individual variability in functional organization. This trajectory is not only consistent with anatomical connectivity but also with the ignition dynamics described in Chapter 1, where LIP emerged as a key node for conscious access.

But the dorsal stream is not the only visual pathway emerging from V1. It is now widely accepted that at least three distinct visual streams originate from this primary area and travel along the cortical surface (Ritchie et al., 2024). The most well-known is the ventral stream (Reddy and Kanwisher, 2006), which curves beneath the cortex and leads to high-level regions involved in abstract visual concepts such as faces and objects. A third stream, often referred to as the lateral visual pathway (Gomez et al., 2019; Rolls et al., 2023), turns around the cortex, passing behind area MT and move towards the Wernicke region in the temporal cortex. What is striking is that these streams appear to follow the same logic as the dorsal one. Each begins in V1 and advances by weaving between transmodal hubs. The ventral stream lies between the middle temporal gyrus and the parahippocampal regions. The lateral stream moves along a corridor between the angular gyrus and the middle temporal gyrus. In both cases, the surrounding transmodal zones belong to the default mode network and seem to act as suppressive boundaries. Our preliminary analyses of functional connectivity to V1 support this view. Negative sources appear consistently in these transmodal regions, matching the pattern seen along the dorsal path These observations suggest a broader organizing principle. Visual streams do not diffuse randomly across the cortex. Instead, they navigate around the core hubs of the default mode network, creating functional pathways that are both shaped and constrained by the surrounding transmodal architecture.

What strongly caught my curiosity is that this stream-like organization, so well described for visual processing, appears to extend to the two other main sensory modalities. At least two distinct streams have been identified originating from the primary somatosensory cortex (Niu et al., 2024; James and Kim, 2010; Reed et al., 2005; Kaas et al., 2011), in a framework similar to the one proposed by Felleman and Van Essen (Felleman and Van Essen, 1991). Remarkably, both of these streams also curve around the core nodes of the default mode network. The first travels along the cingulate cortex (Arienzo et al., 2006; Sun et al., 2023). In doing so, it bypasses the dorsal prefrontal regions of the default mode network. The second moves toward the intraparietal sulcus, avoiding both the angular gyrus and the postero-medial cortex—following almost the exact route of the dorsal visual stream—and ultimately reaches LIP, where somatosensory-selective regions have also been identified (Sereno and Huang, 2006; Huang and Sereno, 2018). A similar pattern is observed for auditory processing, though relatively few studies directly examine the existence of distinct processing streams per se. Still, available evidence supports the notion of a structured progression from primary auditory cortex toward regions implicated in higher-level auditory functions,

such as speech comprehension. From the primary auditory cortex, at least two distinct streams seem to emerge. One follows a lateral trajectory, escaping both the middle temporal gyrus and the angular gyrus (Hamilton et al., 2021), and could be converging towards the Wernicke area. The second is a ventral stream that would travels between the middle temporal cortex and the parahippocampal region, eventually moves forward the fusiform face area (Rauschecker, 2012; Cohen et al., 2016; Hickok and Poeppel, 2004). Two aspects of this organization stand out. First, across all modalities, sensory information appears to travel along surface-based streams that consistently avoid transmodal hubs. Whenever there is space between two default mode network nodes, a sensory stream seems to pass through. Second, these streams do not operate in isolation. They converge. Each one links different sensory modalities but terminates in a common target. The dorsal stream connects visual and somatosensory inputs in LIP (Sereno and Huang, 2006; Huang and Sereno, 2018). The ventral stream brings together auditory and visual inputs in the fusiform face area (Kawase et al., 2005; Awwad Shiekh Hasan et al., 2016). The lateral stream links both auditory and visual signals in Wernicke's area (Pratt et al., 2015; Binder et al., 1997). These three regions—LIP, FFA, and Wernicke—share two important features. First, they all process multimodal information. LIP integrates visual and somatosensory inputs for spatial representation. Wernicke's area supports semantic processing from both auditory and visual language cues. The fusiform face area responds to category-level representations that can be evoked through both sound and vision. In each case, the region can be activated by at least two sensory modalities. Second, all three regions seem to maintain strong long-range connections to other parts of the cortex (Scolari et al., 2015; Szczepanski et al., 2013; Friederici, 2015; Tomasi and Volkow, 2012; Saygin et al., 2012). Given these properties, it seems consistent that they would sit near the top of their respective processing streams. They appear to serve as convergence zones for integrating distributed inputs, and as gateways for broadcasting this integrated information to the highest associative areas of the brain. However, this functional positioning remains hypothetical. While existing evidence supports the view of these regions as integrative hubs with broad cortical reach, a systematic and exhaustive review is currently underway to substantiate this framework with greater anatomical and functional specificity.

This leads us to propose a framework for functional organization in the cortex, grounded in the geometry of the cortical surface. While this framework is currently best supported in the posterior regions of the brain, its principles may apply more broadly. In this view, sensory information does not diffuse randomly across the cortex. Instead, it travels along structured pathways, which we refer to as streams. These streams originate in primary sensory regions and follow reproducible trajectories shaped by both local connectivity and global anatomical constraints. A key feature of this architecture is the consistent presence of transmodal hubs along the cortical surface. These hubs contain key regions of the default mode network and appear to exert a negative influence on the surrounding cortex. Rather than facilitating

the spread of sensory information, they shape its path indirectly, creating corridors through which the streams must pass. At specific locations, these streams converge. The resulting convergence zones, such as LIP, Wernicke's area, and the fusiform face area, receive inputs from different sensory modalities. They combine this information into more abstract or behaviorally relevant representations and transmit it forward to higher-order associative areas. In this sense, they function both as integration sites and as broadcast centers within the larger cortical network. What emerges from this architecture is not a mosaic of isolated modules, but a system of gradients of functional organization underlying functional integration. The geometry of the cortical sheet plays an active role in shaping where information flows, how it is transformed, and how it becomes accessible to distributed cognitive processes. This view contrasts with the influential theory proposed by Smallwood and colleagues (Smallwood et al., 2021), in which cortical organization follows a gradient of abstraction: from concrete sensory processing in primary areas to increasingly abstract representations, culminating in the default mode network (DMN), positioned at the apex of the cortical hierarchy. In their framework, the DMN's functional role emerges passively from its location—furthest from the primary anchors and at the convergence of multiple sensory streams. While I largely adhere to this gradient-based view, our results suggest a crucial departure: instead of converging toward the DMN, sensory streams appear to actively avoid it. Rather than being a passive apex, the DMN seems to shape the architecture itself, exerting suppressive influences and possibly acting as a source of internally generated activity—such as hippocampal replays. In this revised perspective, the DMN is not just a convergence zone but an additional anchor, organizing the flow of information alongside the primary sensory systems.

Conclusion, Where the Stream Begins

What I learnt

> *In the nurturing schema, we celebrate the courage and openness of scientists. Sailing into the unknown again and again takes courage; seeing there something different from expectations, and usually more rich and strange, requires uncommon openness.*
>
> Uri Alon

Over the course of this dissertation—and my scientific journey—I began with a simple intuition: that something happens when we become aware of the song of a bird. My obsession became to trace that moment, to capture the dance of large-scale cortical dynamics that might bridge the physical fabric of the brain and the texture of subjective experience.

In Chapter 1, we focused on conscious access—that pivotal transition when a stimulus crosses the threshold into awareness. By simulating cortical activity across thousands of network architectures, we identified a core mathematical principle: a dynamic bifurcation. What makes this result beautiful to me is its generality. It is not tied to a specific neuron or region, but emerges from the dynamics of the system itself. Even in the presence of noise and local variability, the population exhibits a sharp transition—a shift from rest to a high-activity attractor state. This all-or-nothing dynamic appears not only in individual populations but across associative subnetworks. In particular, we identified a distributed network of associative areas—spanning parietal, temporal, and prefrontal regions—that consistently exhibited bistable dynamics and acted as a candidate for the brain's global workspace. It suggests a generic mechanism for broadcasting information across the cortex. And because only one attractor can dominate at a time, this mechanism introduces competition—perhaps the very process

that shapes the stream of consciousness, with each moment of awareness chasing out the last, like the song of the bird being replaced by the memory of another melody.

In Chapter 2, we shifted focus to the song itself—not its arrival into awareness, but how its content is represented across the cortex. The fractal theory of brain organization has long been challenged, but we pushed its implications further by asking a simple question: How is the most basic piece of visual information—an object's position in the retinotopic space—encoded in the brain? This is just one fragment of our conscious experience—far from complete without the bird's shape, color, or motion—and a substantial portion of the underlying activity may support only preconscious or non-conscious processing of that information. But already, over 18% of the cortex appeared to be involved in encoding this single feature. To make sense of this overwhelming richness, we turned to machine learning. These tools allowed us to reduce the complexity of cortical activation into a compact set of cortical spatial influences. What emerged from this analysis was a picture of broad spatial competition. Rather than fixed, modular maps, we found distributed patterns shaped by interactions across distant cortical territories. Certain key areas exerted consistent influence on the representation of others, turning the cortex into a kind of battleground, where overlapping streams of information compete for expression. Just as Chapter 1 uncovered a temporal competition among neural states, this chapter revealed a spatial competition—a geometry of influence that helps shape how the brain encodes experience. What emerged from this analysis was a broad spatial competition. Key areas exerted consistent influence on the representation of information, turning the cortex into a battleground, where overlapping waves of information compete for expression.

Finally, in Chapter 3, we turned to the underlying mechanisms that might support the spatial competition described in Chapter 2. Specifically, we explored how the cortex—essentially a two-dimensional sheet composed of highly plastic neurons—might self-organize based on the interplay of sensory input and experience-dependent reward signals. This chapter is more speculative, and the results more preliminary. Still, what emerged was a parsimonious model of how the cortex might form a functional blueprint capable of supporting flexible and resilient cognition—integrating diverse sources of information while amplifying what matters most. From this, we proposed a tentative macro-scale architecture for the posterior cortex. In this framework, anchoring points—regions receiving direct thalamic input—serve as gateways for sensory information. The activity they receive sculpts the surrounding cortex, initiating processing streams that traverse the cortical sheet. These streams tend to bypass transmodal hubs containing the Default Mode and Control networks, which are more closely linked to internally generated cognition. Instead, they converge at integration zones—regions like LIP, Wernicke's area, or the fusiform face area (FFA)—that lie at the boundaries between major transmodal hubs. These zones may act as multimodal convergence

points, where information is integrated across sensory domains. From there, it may be broadcast across the cortex via long-range white matter tracts. These integrative regions could be key components of the associative bistable network identified in Chapter 1—part of the larger circuit responsible for conscious access.

What I didn't

Each of these chapters traced a different layer of the same puzzle: how fleeting experiences arise from structured neural activity. Each level reveals a different facet of how the brain might shape experience—not through fixed modules or labeled lines, but through dynamical patterns and spatial interactions. And yet, for all we've uncovered, much remains unresolved. The models are still idealized, the data partial, and the leap from neural geometry to lived experience still vast. The next step is to refine our tools—not just to simulate or decode, but to explain.

In Chapter 1, the simulations were based on the macaque connectome. While useful as a proxy, this limits direct interpretability in the context of human conscious experience. Adapting the framework to human brain anatomy—leveraging high-resolution connectomics and individual variability—will be a crucial next step. Another limitation is the current treatment of the attractor dynamics. So far, the attractor space in our model is binary and content-agnostic: each population has access to just two predefined receptive fields, replicated across the entire cortex. The same input profiles are present in primary visual, somatosensory, and associative areas alike. But real experience is shaped by what is represented—whether it's a face, a touch, or a melody. A crucial next step is to enrich the model's representational repertoire: to embed content-specific attractors whose geometry varies across regions. Here, the spatial decomposition developed in Chapter 2 could act as a constraint—suggesting how different types of information (spatial, semantic, emotional) influence the structure of these attractors, and how selective amplification emerges from the cortical landscape itself. Moreover, attractors need not be fixed. Ongoing experience likely induces a constant reshaping of the ignited subnetwork, with areas joining or leaving the workspace depending on task demands, memory, or internal state. One idea that continues to haunt me is whether the bistable network we identified might act, under certain conditions, as a chaotic attractor—especially during spontaneous, self-generated activity like mind-wandering or dreaming. This remains entirely speculative, but suggests a direction for extending the model toward more fluid, continuous forms of conscious dynamics. Finally, the platform built in Chapter 1 offers opportunities well beyond consciousness. By perturbing these virtual brains—introducing simulated strokes, epileptic dynamics, or pharmacological manipulations—we can begin to ask causal questions that are difficult to explore in vivo. Preliminary steps in this direction have already been taken,

and they open a pathway toward computational phenotyping and even predictive neuromodulation.

Chapter 2 revealed spatial influences shaping how information is distributed across the cortex—but it's important to acknowledge that the analysis was designed to find such influences in the first place. While the results are robust, they should not be interpreted as definitive evidence of causal mechanisms. In the current model, each influence is treated as if it emerges from a single cortical vertex. But in reality, these effects likely arise from groups of vertices—such as the dense thalamic inputs into V1—which collectively shape local dynamics. More importantly, the identified influences may reflect correlations rather than causation. Their primary value is hypothesis-generating: they offer clues about the underlying mechanisms, but those mechanisms must be tested more directly. Chapter 3 represents an initial step in that direction, but far more work is needed to translate the abstract patterns into biological detail. More broadly, the preliminary results we've obtained in other modalities and higher-level domains are promising—but still incomplete. Extending this decomposition to internally driven processes, such as theory of mind, episodic memory, or mental imagery, would be particularly powerful. Such work could help confirm the proposed spatial symmetry between sensory-dominated regions and transmodal hubs like the Default Mode Network. Another important direction is comparative and developmental application. This framework could be applied to other species with known retinotopic maps—such as macaques, mice, or even infants and young children—to investigate how these spatial influences evolve across phylogeny and ontogeny. Finally, the method we've developed is intentionally general. It can be applied to any cortical map—not just functional ones, but also structural features, gene expression patterns, or even contrast maps from neuroimaging studies. Our goal is to make this tool accessible and flexible, so it can support a wide range of scientific questions beyond those explored here.

Finally, Chapter 3 is the most speculative part of this work, offering a tentative synthesis that spans many of the directions I'm actively pursuing. Whether this story will hold up—or how much of it will need to be rewritten—remains to be seen.

The horizon

What is striking about the mechanisms we explored is their generality. They are not specific to humans. Not even to primates. Every mammal has neurons. Every mammalian brain is shaped by the same types of electrical and chemical interactions, and by the same principles of plasticity. The dynamics we uncovered—the conditions for ignition, the spatial competition or the patterns of self-organization—are not unique to our species. They are a product of the physics of neurons, and so they likely apply to any mammalian brain. But the implications might go even further. The ignition dynamics described do not depend on carbon-based biology. They depend on a set

of equations. Any physical system that follows those equations could, in principle, exhibit the same behavior. Similarly, the self-organizing principles are not tied to DNA or synapses. They are the result of competitive interactions under spatial constraints, and could emerge in very different substrates. This leads to a question I cannot stop asking. If evolution on Earth discovered these dynamical properties—if it found ways to build cognitive systems from them—could a similar evolution algorithm elsewhere do the same? Could life on other planets, using very different materials, stumble on the same principles?

A century ago, the idea of life beyond Earth seemed improbable. Today, humanity is committing billions of dollars to the search for signs of life on Mars, Titan, and Enceladus. My intuition tells me—and I'm deeply convinced—that by the end of this century we will know that life has emerged independently at least a couple of times within our very own solar system. Such a discovery would transform our view of the universe, signaling that life is not an earthly quirk but a cosmic commonplace. After all, the Milky Way alone contains hundreds of billions of stars. What drives me deeper still is the uniformity of physical law across the cosmos. Everywhere we look, matter organizes into spiraling galaxies(Figure 3.4). Stars form and shepherd planets into orbit. These structures emerge because, at every scale, the universe's governing equations favor certain patterns—attractors in the dynamics of matter and energy. So I return to my research with a haunting question: the ignition patterns and self-organizing cortical architectures we've studied—could they, too, be attractors of physical law? Is subjective experience itself a natural consequence, replicated wherever the right conditions arise? I may never find answer —but I hope that my journey brings me closer.



Figure 3.4: *Captured by James Webb*

Table of Contents

Résumé substantiel **iii**

Abstract **ix**

Résumé **xii**

Introduction **1**

1 Distributed Dynamics Underlying Conscious Access **8**
1.1 What Do We Mean by Consciousness? 8
1.2 The Physics of Neurons 11
1.3 Bifurcation in a Neural Population 13
1.4 Bifurcation in the Cortex 18
1.5 Introducing Atoum: A Synthetic Cortex for Modeling Awareness . . . 22
1.6 Predicting the Brain 25

2 Spatial Influences on Cortical Sensory Representation **28**
2.1 The Cortical Distribution of Mental Representation 29
2.2 The Tethering Hypothesis 33
2.3 Spatial Component Decomposition 36
2.4 Revealing Structured Spatial Influences on Retinotopic Maps 42
2.5 Open Questions and Future Work 47

3 Self-Organizing Cortex: Mechanisms and Proposed Architecture **49**
3.1 The Formation of Receptive Fields 50
3.2 Global Architecture Emerging from Spatial Competition 55

Conclusion **59**

Bibliography **5**

Bibliography

Albantakis, L., Barbosa, L., Findlay, G., Grasso, M., Haun, A. M., Marshall, W., Mayner, W. G. P., Zaeemzadeh, A., Boly, M., Juel, B. E., Sasai, S., Fujii, K., David, I., Hendren, J., Lang, J. P., and Tononi, G. (2023). Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms. *PLOS Computational Biology*, 19(10):e1011465.

Arienzo, D., Babiloni, C., Ferretti, A., Caulo, M., Del Gratta, C., Tartaro, A., Rossini, P., and Romani, G. (2006). Somatotopy of anterior cingulate cortex (ACC) and supplementary motor area (SMA) for electric stimulation of the median and tibial nerves: An fMRI study. *NeuroImage*, 33(2):700–705.

Awwad Shiekh Hasan, B., Valdes-Sosa, M., Gross, J., and Belin, P. (2016). "Hearing faces and seeing voices": Amodal coding of person identity in the human brain. *Scientific Reports*, 6(1):37494.

Barrett, H. C., Bolyanatz, A., Broesch, T., Cohen, E., Froerer, P., Kanovsky, M., Schug, M. G., and Laurence, S. (2021). Intuitive Dualism and Afterlife Beliefs: A Cross-Cultural Study. *Cognitive Science*, 45(6).

Benson, N. C., Jamison, K. W., Arcaro, M. J., Vu, A. T., Glasser, M. F., Coalson, T. S., Van Essen, D. C., Yacoub, E., Ugurbil, K., Winawer, J., and Kay, K. (2018). The Human Connectome Project 7 Tesla retinotopy dataset: Description and population receptive field analysis. *Journal of Vision*, 18(13):23.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., and Prieto, T. (1997). Human Brain Language Areas Identified by Functional Magnetic Resonance Imaging. *The Journal of Neuroscience*, 17(1):353–362.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2):227–247.

Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., and Tononi, G. (2017). Are the Neural Correlates of Consciousness in the Front or in the Back of the Cerebral Cortex? Clinical and Neuroimaging Evidence. *The Journal of Neuroscience*, 37(40):9603–9613.

Bressler, D. W. and Silver, M. A. (2010). Spatial attention improves reliability of fMRI retinotopic mapping signals in occipital and parietal cortex. *NeuroImage*, 53(2):526–533.

Brodmann, K. (2005). *Brodmann's: Localisation in the Cerebral Cortex*. Springer, New York, NY, 3rd ed edition.

Buckner, R. L. and DiNicola, L. M. (2019). The brain's default network: updated anatomy, physiology and evolving insights. *Nature Reviews Neuroscience*, 20(10):593–608.

Buckner, R. L. and Krienen, F. M. (2013). The evolution of distributed association networks in the human brain. *Trends in Cognitive Sciences*, 17(12):648–665.

Cadwell, C. R., Bhaduri, A., Mostajo-Radji, M. A., Keefe, M. G., and Nowakowski, T. J. (2019). Development and Arealization of the Cerebral Cortex. *Neuron*, 103(6):980–1004.

Callaway, E. M. (2005). Structure and function of parallel pathways in the primate early visual system. *The Journal of Physiology*, 566(1):13–19.

Caporale, N. and Dan, Y. (2008). Spike Timing–Dependent Plasticity: A Hebbian Learning Rule. *Annual Review of Neuroscience*, 31(1):25–46.

Caramazza, A., McCloskey, M., and Green, B. (1981). Naive beliefs in "sophisticated" subjects: misconceptions about trajectories of objects. *Cognition*, 9(2):117–123.

Chalmers, D. J. (1995). Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*.

Chalmers, D. J. (2018). The Meta-Problem of Consciousness. *Journal of Consciousness Studies*.

Citri, A. and Malenka, R. C. (2008). Synaptic Plasticity: Multiple Forms, Functions, and Mechanisms. *Neuropsychopharmacology*, 33(1):18–41.

Clopath, C., Bono, J., and Klatzmann, U. (2015). Excitatory to inhibitory connectivity shaped by synaptic and homeostatic plasticity. *BMC Neuroscience*, 16(1):P126.

Clopath, C., Büsing, L., Vasilaki, E., and Gerstner, W. (2010). Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nature Neuroscience*, 13(3):344–352.

Cohen, Y. E., Bennur, S., Christison-Lagay, K., Gifford, A. M., and Tsunada, J. (2016). Functional Organization of the Ventral Auditory Pathway. In Van Dijk, P., Başkent, D., Gaudrain, E., De Kleine, E., Wagner, A., and Lanting, C., editors, *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*, volume 894, pages 381–388. Springer International Publishing, Cham. Series Title: Advances in Experimental Medicine and Biology.

Dehaene, S., Changeux, J.-P., and Naccache, L. (2011). The Global Neuronal Workspace Model of Conscious Access: From Neuronal Architectures to Clinical Applications. In Dehaene, S. and Christen, Y., editors, *Characterizing Consciousness: From Cognition to the Clinic?*, pages 55–84. Springer Berlin Heidelberg, Berlin, Heidelberg. Series Title: Research and Perspectives in Neurosciences.

Del Cul, A., Baillet, S., and Dehaene, S. (2007). Brain Dynamics Underlying the Nonlinear Threshold for Access to Consciousness. *PLoS Biology*, 5(10):e260.

Descartes, R. (1641). *Meditations on first philosophy: with selections from the objections and replies ; a Latin-English edition.* Cambridge University Pr, Cambridge, England.

Descartes, R. (1984). *The philosophical writings of Descartes.* Cambridge University Press, Cambridge [Cambridgeshire] ; New York.

Destexhe, A., Mainen, Z. F., and Sejnowski, T. J. (1994). Synthesis of models for excitable membranes, synaptic transmission and neuromodulation using a common kinetic formalism. *Journal of Computational Neuroscience*, 1(3):195–230.

Doerig, A., Kietzmann, T. C., Allen, E., Wu, Y., Naselaris, T., Kay, K., and Charest, I. (2025). High-level visual representations in the human brain are aligned with large language models. *Nature Machine Intelligence*, 7(8):1220–1234.

Doerig, A., Schurger, A., Hess, K., and Herzog, M. H. (2019). The unfolding argument: Why IIT and other causal structure theories cannot explain consciousness. *Consciousness and Cognition*, 72:49–59.

Dohmatob, E., Mensch, A., Varoquaux, G., and Thirion, B. (2016). Learning brain regions via large-scale online structured sparse dictionary learning. *NIPS*.

Dumoulin, S. O. and Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, 39(2):647–660.

Ercsey-Ravasz, M., Markov, N., Lamy, C., Van Essen, D., Knoblauch, K., Toroczkai, Z., and Kennedy, H. (2013). A Predictive Network Model of Cerebral Cortical Connectivity Based on a Distance Rule. *Neuron*, 80(1):184–197. Publisher: Elsevier BV.

Felleman, D. J. and Van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, 1(1):1–47.

Fleming, S. M. and Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594):1338–1349.

Forstmann, M. and Burgmer, P. (2015). Adults are intuitive mind-body dualists. *Journal of Experimental Psychology: General*, 144(1):222–235.

Fox, R. F. (1997). Stochastic versions of the Hodgkin-Huxley equations. *Biophysical Journal*, 72.

Friederici, A. D. (2015). White-matter pathways for speech and language processing. In *Handbook of Clinical Neurology*, volume 129, pages 177–186. Elsevier.

Froudist-Walsh, S., Bliss, D. P., Ding, X., Rapan, L., Niu, M., Knoblauch, K., Zilles, K., Kennedy, H., Palomero-Gallagher, N., and Wang, X.-J. (2021). A dopamine gradient controls access to distributed working memory in the large-scale monkey cortex. *Neuron*, 109(21):3500–3520.e13.

Froudist-Walsh, S., Xu, T., Niu, M., Rapan, L., Zhao, L., Margulies, D. S., Zilles, K., Wang, X.-J., and Palomero-Gallagher, N. (2023). Gradients of neurotransmitter receptor expression in the macaque cortex. *Nature Neuroscience*, 26(7):1281–1294.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., and Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178.

Gomez, J., Drain, A., Jeska, B., Natu, V. S., Barnett, M., and Grill-Spector, K. (2019). Development of population receptive fields in the lateral visual stream improves spatial coding amid stable structural-functional coupling. *NeuroImage*, 188:59–69.

Grill-Spector, K., Kourtzi, Z., and Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, 41(10-11):1409–1422.

Grill-Spector, K. and Malach, R. (2004). THE HUMAN VISUAL CORTEX. *Annual Review of Neuroscience*, 27(1):649–677.

Hackett, T. A. (2011). Information flow in the auditory cortical network. *Hearing Research*, 271(1-2):133–146.

Haenelt, D., Trampel, R., Nasr, S., Polimeni, J. R., Tootell, R. B., Sereno, M. I., Pine, K. J., Edwards, L. J., Helbling, S., and Weiskopf, N. (2023). High-resolution quantitative and functional MRI indicate lower myelination of thin and thick stripes in human secondary visual cortex. *eLife*, 12:e78756.

Hamilton, L. S., Oganian, Y., Hall, J., and Chang, E. F. (2021). Parallel and distributed encoding of speech across human auditory cortex. *Cell*, 184(18):4626–4639.e13.

Hickok, G. and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2):67–99.

Himmelberg, M. M., Kurzawski, J. W., Benson, N. C., Pelli, D. G., Carrasco, M., and Winawer, J. (2021). Cross-dataset reproducibility of human retinotopic maps. *NeuroImage*, 244:118609.

Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4):500–544.

Huang, R.-S. and Sereno, M. I. (2018). Multisensory and sensorimotor maps. In *Handbook of Clinical Neurology*, volume 151, pages 141–161. Elsevier.

Huntenburg, J. M., Bazin, P.-L., and Margulies, D. S. (2018). Large-Scale Gradients in Human Cortical Organization. *Trends in Cognitive Sciences*, 22(1):21–31.

IIT-Concerned, Fleming, S. M., Frith, C. D., Goodale, M., Lau, H., LeDoux, J. E., Lee, A. L. F., Michel, M., Owen, A. M., Peters, M. A. K., and Slagter, H. A. (2023). The Integrated Information Theory of Consciousness as Pseudoscience.

Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3(3):318–356.

James, T. W. and Kim, S. (2010). Dorsal and Ventral Cortical Pathways for Visuo-haptic Shape Integration Revealed Using fMRI. In Kaiser, J. and Naumer, M. J., editors, *Multisensory Object Perception in the Primate Brain*, pages 231–250. Springer New York, New York, NY.

James, W. (1890). The Principles of Psychology. *The Principles of Psychology*.

Jay, T. M. (2003). Dopamine: a potential substrate for synaptic plasticity and memory mechanisms. *Progress in Neurobiology*, 69(6):375–390.

Jirsa, V., Wang, H., Triebkorn, P., Hashemi, M., Jha, J., Gonzalez-Martinez, J., Guye, M., Makhalova, J., and Bartolomei, F. (2023). Personalised virtual brain models in epilepsy. *The Lancet Neurology*, 22(5):443–454.

Joglekar, M. R., Mejias, J. F., Yang, G. R., and Wang, X.-J. (2018). Inter-areal Balanced Amplification Enhances Signal Propagation in a Large-Scale Circuit Model of the Primate Cortex. *Neuron*, 98(1):222–234.e8.

Kaas, J. H. (1993). The functional organization of somatosensory cortex in primates. *Annals of Anatomy - Anatomischer Anzeiger*, 175(6):509–518.

Kaas, J. H., Gharbawie, O. A., and Stepniewska, I. (2011). The Organization and Evolution of Dorsal Stream Multisensory Motor Pathways in Primates. *Frontiers in Neuroanatomy*, 5.

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience*.

Katz, B. and Miledi, R. (1967). A study of synaptic transmission in the absence of nerve impulses. *The Journal of Physiology*, 192(2):407–436.

Kawase, T., Yamaguchi, K., Ogawa, T., Suzuki, K.-i., Suzuki, M., Itoh, M., Kobayashi, T., and Fujii, T. (2005). Recruitment of fusiform face area associated with listening to degraded speech sounds in auditory–visual speech perception: a PET study. *Neuroscience Letters*, 382(3):254–258.

Kim, Y., Yang, G. R., Pradhan, K., Venkataraju, K. U., Bota, M., García Del Molino, L. C., Fitzgerald, G., Ram, K., He, M., Levine, J. M., Mitra, P., Huang, Z. J., Wang, X.-J., and Osten, P. (2017). Brain-wide Maps Reveal Stereotyped Cell-Type-Based Cortical Architecture and Subcortical Sexual Dimorphism. *Cell*, 171(2):456–469.e22.

Klatzmann, U., Froudist-Walsh, S., Bliss, D. P., Theodoni, P., Mejías, J., Niu, M., Rapan, L., Palomero-Gallagher, N., Sergent, C., Dehaene, S., and Wang, X.-J. (2025). A dynamic bifurcation mechanism explains cortex-wide neural correlates of conscious access. *Cell Reports*, 44(3):115372.

Kuhlmeier, V. A., Bloom, P., and Wynn, K. (2004). Do 5-month-old infants see humans as material objects? *Cognition*, 94(1):95–103.

Lamme, V. (2015). The Crack of DawnThe Crack of Dawn: Perceptual Functions and Neural Mechanisms that Mark the Transition from Unconscious Processing to Conscious Vision: Perceptual Functions and Neural Mechanisms that Mark the Transition from Unconscious Processing to Conscious Vision. *Open MIND*. Publisher: Theoretical Philosophy/MIND Group – JGU Mainz.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Luppi, A. I., Vohryzek, J., Kringelbach, M. L., Mediano, P. A. M., Craig, M. M., Adapa, R., Carhart-Harris, R. L., Roseman, L., Pappas, I., Peattie, A. R. D., Manktelow, A. E., Sahakian, B. J., Finoia, P., Williams, G. B., Allanson, J., Pickard, J. D., Menon, D. K., Atasoy, S., and Stamatakis, E. A. (2023). Distributed harmonic patterns of structure-function dependence orchestrate human consciousness. *Communications Biology*, 6(1):117.

Mackey, W. E., Winawer, J., and Curtis, C. E. (2017). Visual field map clusters in human frontoparietal cortex. *eLife*, 6:e22974.

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., Bezgin, G., Eickhoff, S. B., Castellanos, F. X., Petrides, M., Jefferies, E., and Smallwood, J. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences*, 113(44):12574–12579.

Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., Misery, P., Falchier, A., Quilodran, R., Gariel, M. A., Sallet, J., Gamanut, R., Huissoud, C., Clavagnier, S., Giroud, P., Sappey-Marinier, D., Barone, P., Dehay, C., Toroczkai, Z., Knoblauch, K., Van Essen, D. C., and Kennedy, H. (2014). A Weighted and Directed Interareal Connectivity Matrix for Macaque Cerebral Cortex. *Cerebral Cortex*, 24(1):17–36.

Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs. *Science*, 275(5297):213–215.

Martini, F. J., Guillamón-Vivancos, T., Moreno-Juan, V., Valdeolmillos, M., and López-Bendito, G. (2021). Spontaneous activity in developing thalamic and cortical sensory networks. *Neuron*, 109(16):2519–2534.

Mashour, G. A., Roelfsema, P., Changeux, J.-P., and Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron*, 105(5):776–798.

McCloskey, M. (1983). Naive Theories of Motion. In *Mental Models*, pages 299–324. Dedre Gentner; Albert L. Stevens, Hillsdale, NJ.

McCloskey, M., Caramazza, A., and Green, B. (1980). Curvilinear Motion in the Absence of External Forces: Naïve Beliefs About the Motion of Objects. *Science*, 210(4474):1139–1141.

Mejías, J. F. and Wang, X.-J. (2022). Mechanisms of distributed working memory in a large-scale network of macaque neocortex. *eLife*, 11:e72136.

Mendonça, D., Curado, M., and Gouveia, S. S., editors (2021). *The philosophy and science of predictive processing*. Bloomsbury Academic, New York, NY.

Miller, K. D., Keller, J. B., and Stryker, M. P. (1989). Ocular Dominance Column Development: Analysis and Simulation. *Science*, 245(4918):605–615.

Monod, J. (1970). *Le Hasard et la Nécessité: Essai sur la philosophie naturelle de la biologie moderne*. Éditions du Seuil, Paris.

Muller, L., Chavane, F., Reynolds, J., and Sejnowski, T. J. (2018). Cortical travelling waves: mechanisms and computational principles. *Nature Reviews Neuroscience*, 19(5):255–268.

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D., and Wang, X.-J. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience*, 17(12):1661–1663.

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, 83(4):435.

Newton, I. (1687). *The Mathematical Principles of Natural Philosophy*. Daniel Adee, New York. English translation of \textit{Philosophi{\ae} Naturalis Principia Mathematica}.

Niu, M., Rapan, L., Froudist-Walsh, S., Zhao, L., Funck, T., Amunts, K., and Palomero-Gallagher, N. (2024). Multimodal mapping of macaque monkey somatosensory cortex. *Progress in Neurobiology*, 239:102633.

Normandin, S. and Wolfe, C. T., editors (2013). *Vitalism and the Scientific Image in Post-Enlightenment Life Science, 1800-2010*, volume 2 of *History, Philosophy and Theory of the Life Sciences*. Springer Netherlands, Dordrecht.

Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLOS Computational Biology*, 10(5):25.

Pang, J. C., Aquino, K. M., Oldehinkel, M., Robinson, P. A., Fulcher, B. D., Breakspear, M., and Fornito, A. (2023). Geometric constraints on human brain function. *Nature*, 618(7965):566–574.

Paquola, C., Vos De Wael, R., Wagstyl, K., Bethlehem, R. A. I., Hong, S.-J., Seidlitz, J., Bullmore, E. T., Evans, A. C., Misic, B., Margulies, D. S., Smallwood, J., and Bernhardt, B. C. (2019). Microstructural and functional gradients are increasingly dissociated in transmodal cortices. *PLOS Biology*, 17(5):e3000284.

Pascucci, D., Menétrey, M. Q., Passarotto, E., Luo, J., Paramento, M., and Rubega, M. (2025). EEG brain waves and alpha rhythms: Past, current and future direction. *Neuroscience & Biobehavioral Reviews*, 176:106288.

Phelps, E. A. and LeDoux, J. E. (2005). Contributions of the Amygdala to Emotion Processing: From Animal Models to Human Behavior. *Neuron*, 48(2):175–187.

Pratt, H., Bleich, N., and Mittelman, N. (2015). Spatio-temporal distribution of brain activity associated with audio-visually congruent and incongruent speech and the McGurk Effect. *Brain and Behavior*, 5(11):e00407.

Puce, A., Allison, T., Bentin, S., Gore, J. C., and McCarthy, G. (1998). Temporal Cortex Activation in Humans Viewing Eye and Mouth Movements. *The Journal of Neuroscience*, 18(6):2188–2199.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045):1102–1107.

Rakic, P., Ayoub, A. E., Breunig, J. J., and Dominguez, M. H. (2009). Decision by division: making cortical maps. *Trends in Neurosciences*, 32(5):291–301.

Rauschecker, J. P. (2012). Ventral and dorsal streams in the evolution of speech and language. *Frontiers in Evolutionary Neuroscience*, 4.

Reddy, L. and Kanwisher, N. (2006). Coding of visual objects in the ventral stream. *Current Opinion in Neurobiology*, 16(4):408–414.

Reed, C. L., Klatzky, R. L., and Halgren, E. (2005). What vs. where in touch: an fMRI study. *NeuroImage*, 25(3):718–726.

Reynolds, J. N. and Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks*, 15(4-6):507–521.

Ritchie, J. B., Montesinos, S., and Carter, M. J. (2024). What Is a Visual Stream? *Journal of Cognitive Neuroscience*, 36(12):2627–2638.

Rolls, E. T., Deco, G., Huang, C.-C., and Feng, J. (2023). Multiple cortical visual streams in humans. *Cerebral Cortex*, 33(7):3319–3349.

Rumelhart, D. E., Hintont, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*.

Saygin, Z. M., Osher, D. E., Koldewyn, K., Reynolds, G., Gabrieli, J. D. E., and Saxe, R. R. (2012). Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nature Neuroscience*, 15(2):321–327.

Scolari, M., Seidl-Rathkopf, K. N., and Kastner, S. (2015). Functions of the human frontoparietal attention network: Evidence from neuroimaging. *Current Opinion in Behavioral Sciences*, 1:32–39.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3):417–424.

Sereno, M. I. and Huang, R.-S. (2006). A human parietal face area contains aligned head-centered visual and tactile maps. *Nature Neuroscience*, 9(10):1337–1343.

Sergent, C., Baillet, S., and Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, 8(10):1391–1400.

Sergent, C., Corazzol, M., Labouret, G., Stockart, F., Wexler, M., King, J.-R., Meyniel, F., and Pressnitzer, D. (2021). Bifurcation in brain dynamics reveals a signature of conscious processing independent of report. *Nature Communications*, 12(1):1149.

Simons, D. J. and Chabris, C. F. (1999). Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events. *Perception*, 28(9):1059–1074.

Smallwood, J., Bernhardt, B. C., Leech, R., Bzdok, D., Jefferies, E., and Margulies, D. S. (2021). The default mode network in cognition: a topographical perspective. *Nature Reviews Neuroscience*, 22(8):503–513.

Song, Y., Lukasiewicz, T., Xu, Z., and Bogacz, R. (2021). Can the Brain Do Backpropagation? —Exact Implementation of Backpropagation in Predictive Coding Networks.

Sood, M. R. and Sereno, M. I. (2016). Areas activated during naturalistic reading comprehension overlap topological visual, auditory, and somatotomotor maps: Reading Overlaps Sensorimotor Maps. *Human Brain Mapping*, 37(8):2784–2810.

Sorabji, R. (2014). *Moral conscience through the ages: fifth century BCE to the present*. Oxford university press, Oxford.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99(2):195–231.

Strogatz, S. H. (2015). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview Press, a member of the Perseus Books Group, Boulder, CO, second edition edition.

Sun, G., McCartin, M., Liu, W., Zhang, Q., Kenefati, G., Chen, Z. S., and Wang, J. (2023). Temporal pain processing in the primary somatosensory cortex and anterior cingulate cortex. *Molecular Brain*, 16(1):3.

Szczepanski, S. M., Pinsk, M. A., Douglas, M. M., Kastner, S., and Saalmann, Y. B. (2013). Functional and structural architecture of the human dorsal frontoparietal attention network. *Proceedings of the National Academy of Sciences*, 110(39):15806–15811.

Sękowski, K. and Rorot, W. (2022). Intuition-Driven Navigation of the Hard Problem of Consciousness. *Review of Philosophy and Psychology*, 13(1):239–255.

Tomasi, D. and Volkow, N. D. (2012). Resting functional connectivity of language networks: characterization and reproducibility. *Molecular Psychiatry*, 17(8):841–854.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5(1):42.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., and Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, 80:62–79.

van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., and Roelfsema, P. R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science*, 360(6388):537–542.

Varoquaux, G., Gramfort, A., Pedregosa, F., Michel, V., and Thirion, B. (2011). Multi-subject Dictionary Learning to Segment an Atlas of Brain Spontaneous Activity. In Székely, G. and Hahn, H. K., editors, *Information Processing in Medical Imaging*, volume 6801, pages 562–573. Springer Berlin Heidelberg, Berlin, Heidelberg. Series Title: Lecture Notes in Computer Science.

Vohryzek, J., Sanz-Perl, Y., Kringelbach, M. L., and Deco, G. (2025). Human brain dynamics are shaped by rare long-range connections over and above cortical geometry. *Proceedings of the National Academy of Sciences*, 122(1):e2415102122.

Vuilleumier, P., Armony, J. L., Driver, J., and Dolan, R. J. (2001). Effects of Attention and Emotion on Face Processing in the Human Brain. *Neuron*, 30(3):829–841.

Wandell, B. A., Dumoulin, S. O., and Brewer, A. A. (2007). Visual Field Maps in Human Cortex. *Neuron*, 56(2):366–383.

Washington, W. M. and Parkinson, C. L. (2005). *An introduction to three-dimensional climate modeling*. Univ. Science Books, Sausalito, Calif, 2. ed edition.

Watson, J. D. and Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 171(4356):737–738.

Wei, W., Benn, R. A., Scholz, R., Shevchenko, V., Klatzmann, U., Alberti, F., Chiou, R., Wassermann, D., Vanderwal, T., Smallwood, J., and Margulies, D. S. (2024). A function-based mapping of sensory integration along the cortical hierarchy. *Communications Biology*, 7(1):1593.

Whittington, J. C. and Bogacz, R. (2019). Theories of Error Back-Propagation in the Brain. *Trends in Cognitive Sciences*, 23(3):235–250.

Witter, M. P. and Amaral, D. G. (1991). Entorhinal cortex of the monkey: V. Projections to the dentate gyrus, hippocampus, and subicular complex. *Journal of Comparative Neurology*, 307(3):437–459.

Wong, K.-F. (2006). A Recurrent Network Mechanism of Time Integration in Perceptual Decisions. *Journal of Neuroscience*, 26(4):1314–1328.

Wu, J., Yan, T., Zhang, Z., Jin, F., and Guo, Q. (2012). Retinotopic mapping of the peripheral visual field to human visual cortex by functional magnetic resonance imaging. *Human Brain Mapping*, 33(7):1727–1740.

Zilles, K. and Amunts, K. (2010). Centenary of Brodmann's map — conception and fate. *Nature Reviews Neuroscience*, 11(2):139–145.

Zorumski, C. F., Izumi, Y., and Mennerick, S. (2016). Ketamine: NMDA Receptors and Beyond. *The Journal of Neuroscience*, 36(44):11158–11164.

Annex 1

A dynamic bifurcation mechanism explains cortex-wide neural correlates of conscious access

Annex 2

Spatial layout of visual specialization is shaped by competing default mode and sensory networks